HIGHWAY RESEARCH RECORD

Number 290

Nuclear Testing

Construction Variability

Materials Control and Acceptance

6 Reports

Subject Area

32	Cement and Concrete
33	Construction
34	General Materials
35	Mineral Aggregates
62	Foundations (Soils)

HIGHWAY RESEARCH BOARD

DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

Washington, D. C., 1969 Publication 1668

Price: $2.40

Available from

Highway Research Board
National Academy of Sciences
2101 Constitution Avenue
Washington, D.C. 20418

Department of Materials and Construction

R. L. Peyton, Chairman
State Highway Commission of Kansas, Topeka

HIGHWAY RESEARCH BOARD STAFF

R. E. Bollen and W. B. Gunderman

GENERAL MATERIALS DIVISION

John L. Beaton, Chairman
California Division of Highways, Sacramento

COMMITTEE ON MINERAL AGGREGATES
(As of December 31, 1968)

F. E. Legg, Jr., Chairman
University of Michigan, Ann Arbor

William P. Chamberlin
J. T. Corkill
Karl H. Dunn
R. D. Gaynor
Joseph E. Gray
Hans I. Hansen
Eugene Y. Huang
Donald R. Lamb
Thomas D. Larson
W. B. Ledbetter
D. W. Lewis
J. F. McLaughlin
Kenneth A. Nelson
Robert G. Pike
F. A. Renninger
Robert L. Schuster
R. D. Shumway
Norman G. Smith
Travis W. Smith
G. W. Steele
Richard D. Walker
E. A. Whitehurst
D. O. Woolf

EVALUATIONS, SYSTEMS AND PROCEDURES DIVISION

John F. McLaughlin, Chairman
Purdue University, Lafayette, Indiana

COMMITTEE ON NUCLEAR PRINCIPLES AND APPLICATIONS
(As of December 31, 1968)

Donald R. Lamb, Chairman
University of Wyoming, Laramie

D. W. Anderson
Percy L. Blackwell
Richard E. Borup
Wayne R. Brown
Jack Dempsey
Wilbur Dunphy, Jr.
C. Page Fisher
Robin P. Gardner
Hamilton Gray
George L. Hinueber
C. S. Hughes
Harry D. Richardson
Rolands L. Rizenbergs
Arnold E. Schwartz
Ernest T. Selig
Preston C. Smith
John Toman
William G. Weber, Jr.
Thomas G. Williamson
Valentin Worona
Eldon J. Yoder

COMMITTEE ON QUALITY ASSURANCE AND ACCEPTANCE PROCEDURES
(As of December 31, 1968)

Leo D. Sandvig, Chairman
Pennsylvania Department of Highways, Harrisburg

Robert M. Nicotera, Secretary
Pennsylvania Department of Highways, Harrisburg

Verdi Adam
John L. Beaton
Robert N. Blakeslee
D. L. Bloem
Richard L. Davis
Clarence E. DeYoung
Jack H. Dillard
Richard E. Forrestel
John D. Geesaman
William R. Hellwig
John R. Kohnke
F. E. Legg, Jr.
George W. McAlpin
John F. McLaughlin
Thurmul F. McMahon
Lionel T. Murray
Peter Smith
G. W. Steele
Jack E. Stephens
David G. Tunnicliff
Warren B. Warden

Department of Soils, Geology and Foundations

Eldon J. Yoder, Chairman
Purdue University, Lafayette, Indiana

Chester McDowell, Vice Chairman
Texas Highway Department, Austin

HIGHWAY RESEARCH BOARD STAFF

J. W. Guinnee

DIVISION C

O. L. Lund, Chairman
Nebraska Department of Roads, Lincoln

L. F. Erickson, Vice Chairman
Idaho Department of Highways, Boise

COMMITTEE ON SOIL AND ROCK PROPERTIES
(As of December 31, 1968)

G. A. Leonards, Chairman
Purdue University, Lafayette, Indiana

George B. Clark
Robert W. Cunny
E. D'Appolonia
Robert C. Deen
Victor Dolmage
Wilbur I. Duvall
Austin H. Emery
Charles L. Emery
Emmericus C. W. A. Geuze
Bernard B. Gordon
James P. Gould
Stanley J. Johnson
Ernest Jonas
Robert L. Kondner
Charles C. Ladd
Thurmul F. McMahon
Victor Milligan
Roy E. Olson
William H. Perloff
Shailer S. Philbrick
Gregory P. Tschebotarioff
William G. Weber, Jr.
T. H. Wu

Foreword

Five of the six papers in this RECORD are concerned with the all-important matter of construction control and compliance to specifications of construction materials. This control may go by one of many names, from the simple "testing" to the more sophisticated-sounding "quality assurance procedures." But whatever the title, engineers agree that this is an area of critical concern and one in which information is sorely needed. In this RECORD, construction and materials engineers and specification writers will find practical, useful information on subjects ranging from how to minimize errors in the use of nuclear gages for measuring soil density and moisture content to acceptance procedures for concrete.

In the paper by Gardner, the sources of error associated with nuclear gages are identified and several approaches are described whereby the sources of error may be minimized. The potential success of each approach is evaluated.

Variability (its measurement), minimization (its implications in specifications), acceptance testing, and quality assurance provide the common threads tying together the papers by Williamson, Jorgenson, Mills and Fletcher, and Mathews and Metcalf.

Williamson attempts to determine the extent of compaction variability present in fill construction and to identify the causes of variation. After studying several different field tests for measurement of in-place density and moisture content (including nuclear gages), he reports finding variability to be widespread and proposes a technique whereby decisions on overall compaction quality can be made.

Variability in compacted embankments was measured by Jorgensen. His major conclusions were that the variability in percent compaction is large; the average percent compaction was very near the required minimum; and the nuclear gage in the direct transmission position is a much more reliable indicator of field density than when in a backscatter position, and is slightly more reliable than the conventional water-balloon tests.

Development of a statistical procedure for control and acceptance of gradations of aggregates was attempted by Mills and Fletcher. Their report contains valuable information derived from random sampling at the aggregate source, from stockpiles at the project sites, and from the aggregate as used in the work. The authors found that samples taken at the source generally conformed to specifications and exhibited low variability, but that as one proceeded to the on-site stockpiles and as-used samples, variability increased and stockpile samples were frequently out of specification limits. They applied the results of the research to the development of models for specifications.

Mathews and Metcalf studied the statistical implications of a United Kingdom specification of strength of concrete for highway structures and the quality of materials accepted to see if the specification could be improved. They concluded, most interestingly, that the quality of work actually produced was good but was little affected by specification requirements. They speculated that the good quality resulted from the unwillingness of contractors to risk the economic consequences of even a low failure rate.

In a paper unrelated to acceptance testing but concerned with a subject of critical importance to pavement stability, Dunn reports the results of a field study made to determine the amount of degradation that occurred in untreated aggregate base courses by manipulation, compaction, and service exposure. He relates his investigation to previous work, and presents a substantial amount of data to indicate that, for the materials studied (hard and soft crushed dolomite and dolomitic and igneous gravels), the greatest degradation occurred during manipulation and compaction. Attrition during service exposure was not a primary factor, and there was no relationship between degradation and aggregate type or physical properties.

—J. F. McLaughlin

Contents

Minimizing Nuclear Soil Density and Moisture Content Gage Errors

ROBIN P. GARDNER, North Carolina State University, Raleigh

The sources of error in the nuclear soil density gages are identified as sensitivity to variations in sample composition, poor calibration technique, and sensitivity to surface heterogeneities. The errors associated with the nuclear moisture content gages are identified as sensitivity to soil composition, sensitivity to soil density, and poor calibration technique. Several approaches are described and evaluated for minimizing these sources of error, including mathematical analyses of the nuclear gaging principles, the calibration model method, and the dual-gage principle for nuclear density gages.

•GAMMA-RAY scatter and neutron moderation gages have been in use for the measurement of soil density and moisture content since about 1950 (1). The immediate obvious advantages of the nuclear methods were nondestructiveness, measurement speed, and good reproducibility. However, when the nuclear gages were put to use in the field and compared to the existing gravimetric methods, discrepancies appeared and the question of the accuracy of these devices arose. To resolve this question, the Highway Research Board formed the Committee on Nuclear Principles and Applications, and NCHRP Project 10-5 (Density and Moisture Content Measurement by Nuclear Methods) was initiated. This paper describes the work that has been done on Project 10-5, the work in progress on Project 10-5A (an extension of Project 10-5), and other pertinent work in this area that has been initiated since the beginning of Project 10-5.

SOURCES OF NUCLEAR GAGE ERROR

The purpose of the initial phase of work on Project 10-5 (2) was to evaluate the nuclear gages for measuring soil density and moisture content in relation to the conventional gravimetric techniques. It was concluded that the nuclear gage results were more reproducible and potentially more accurate if the identified sources of error could be minimized. Sources of error identified for the neutron moisture content gages were sensitivity to soil density, sensitivity to soil composition, and poor calibration techniques.

The primary source of error for the gamma-ray density gage was sensitivity to composition. The calibration problem stems from the composition sensitivity of the gages and is compounded by difficulties in preparing stable, homogeneous samples of soil for laboratory calibration, or by inaccuracies in the gravimetric density measurement techniques when field calibrations are used. Likewise, the primary sources of error for the neutron moisture content gage were sensitivity to density and composition. Therefore, the calibration problem is essentially the same for the neutron moisture content gages as for the gamma-ray density gages.

The Virginia Correlation and Conference (3) sponsored by the HRB Committee on Nuclear Principles and Applications provided valuable quantitative information on the

---

Paper sponsored by Committee on Nuclear Principles and Applications and presented at the 48th Annual Meeting.

sensitivity of the gamma-ray density gages to soil composition. At the conference laboratory, samples of known density, moisture (hydrogen) content, and composition were available for calibrating gages brought by interested users and gage manufacturers. After the gages were calibrated on these laboratory samples, they were used in the field at a prepared test site composed of five typical Virginia construction soils. The average standard error reported for all backscatter-type gamma-ray density gages on the five laboratory samples was ± 11.0 pcf (pounds per cubic foot), whereas the error for transmission-type gages was ± 7.53 pcf. The average standard error reported for all neutron moisture content gages on the four laboratory samples was ± 1.14 pcf water. These standard errors were determined by fitting the gage responses by a least-squares method to straight-line functions of density or moisture content.

MATHEMATICAL ANALYSES OF THE NUCLEAR GAGES

To quantitatively evaluate the extent of the identified sources of error, mathematical analyses have been made of both the gamma-ray and neutron gages. A brief account of these analyses, some analyses made by other workers, and the conclusions that can be made based on these analyses are given here.

Analyses of the Gamma-Ray Soil Density Gages

In the initial phase of Project 10-5, a simple single-scatter mathematical model of the gamma-ray density gages was developed and tested. The simplifying assumptions were (a) that the gage geometry was identical to that of the depth-type gages so that the problem was only two-dimensional, and (b) that the total gage response was directly proportional to the gamma rays emitted by the source and scattered once by the surrounding soil directly into the detector. Using these two assumptions, the response of the gage could be given as the double integral of the product of four separable probabilities:

$$R = \int_{\phi} \int_{r} \frac{dP_1}{d\phi} \frac{dP_2}{dr} P_3 P_4 \, d\phi \, dr \tag{1}$$

where R is the gage response; ϕ is the angle between the line connecting the source and detector and the direction of the gamma ray being considered; r is the distance from the source to the scattering point; P_1 is the probability that a gamma ray will move in a direction between ϕ and $\phi + d\phi$ and will reach distance r without being scattered or absorbed; P_2 is the probability that a gamma ray described by P_1 will scatter between r and r + dr so that it travels in the direction of the detector; P_3 is the probability that a gamma ray described by P_1 and P_2 reaches the detector from the scattering point without being scattered or absorbed; and P_4 is the detector efficiency for the gamma ray described by P_1, P_2, and P_3.

The four probabilities P_1, P_2, P_3, and P_4 are complicated functions of the original gamma-ray energy, the particular gamma-ray path being considered, and the type of detector (2).

The double integral of Eq. 1 was put into finite difference form and programmed for solution on a digital computer. Solutions of this equation did not reproduce experimental results performed with a prototype depth-type gage on prepared laboratory samples. It was concluded that the second simplifying assumption was not valid, and that the response caused by multiple-scattering events would have to be included in the analysis. This was accomplished by assuming that buildup factors derived from the results reported by Goldstein and Wilkins (4) could be superimposed on each gamma-ray path. The new mathematical model including this buildup factor can be written as

$$R = \int_{\phi} \int_{r} \frac{dP_1}{d\phi} \frac{dP_2}{dr} P_3 P_4 P_5 \, d\phi \, dr \tag{2}$$

where P_5 is the buildup factor for a particular gamma-ray path.

Solutions to Eq. 2 did reproduce experimental results performed with a prototype depth-type gage on laboratory standards. After this verification of the mathematical model given in Eq. 2, it was used extensively to quantitatively study the effect of certain gage parameters on the sensitivity of the gages to soil composition. Among the conclusions reached were that (a) the composition sensitivity of a given gage was essentially independent of the sample density; and (b) composition sensitivity is affected by the gage housing material and thickness, the detector efficiency vs gamma-ray energy relationship, the source-to-detector distance, and the source collimation angles. Quantitative predictions of these effects are given by Ballard and Gardner (2).

A recent series of papers by Taylor and Kansara (5, 6, 7) reports on a single-scatter model of the gamma-ray gages. These authors were apparently unaware of the previous work just discussed. Unfortunately, they concluded that a single-scatter model was sufficient to describe the gamma-ray scatter technique. Therefore, the subsequent conclusions drawn by these authors are questionable.

The comprehensive multiple-scatter model described in this section served to investigate the practical importance of the possible sources of error that were identified. The model was then used to study the effect of varying all possible gage design parameters on these sources of error. The results of these studies indicated that changes in any single gage design parameter tended to minimize the error caused by variations in soil composition only at the expense of increasing the error caused by surface heterogeneities. For example, suppose that very low angle scattering of gamma rays is accentuated by source collimation so that, on the average, higher gamma-ray energies are detected. Then the effect of soil composition variations is minimized because the photoelectric absorption effect is minimized because of the higher average energy of the detected gamma rays, but the effect of surface heterogeneities is increased because the effective sample depth of the gage is reduced and the average relative path length through the surface heterogeneity is increased. This conflicting effect of varying the design parameters indicated that improvements in gage design would not be able to satisfactorily minimize the two sources of error caused by soil composition and surface heterogeneities. Moreover, even if gage design improvements could satisfactorily minimize these two sources of error, this would not solve the problem of using the many existing gages with acceptable error. For these reasons the calibration model approach described later was pursued.

Analyses of the Neutron Soil Moisture Content Gages

Attempts to perform a detailed mathematical analysis of the neutron soil moisture content gages have not been as successful as those for the gamma-ray soil density gages. This is primarily because neutron transport is a more complex and difficult phenomenon than gamma-ray transport. An excellent early study by Semmler (8) gives several possible mathematical approaches that include the use of various forms of one- and two-group neutron diffusion models. This study served as the primary basis for selection of a two-group neutron diffusion model with a spherical cavity of adjustable radius such as the model used in the studies made in the second phase of Project 10-5 (9). This model is given by

$$R = \frac{K_1}{\Sigma a\,(L_1 + K_2)\,(L_2 + K_2)\,(L_1 + L_2)} + K_3 \tag{3}$$

where R is the gage response; Σa is the macroscopic thermal absorption probability of the sample; L_1 is the diffusion length of fast neutrons in the sample; L_2 is the diffusion length of thermal neutrons in the sample; K_1 is a factor that depends on source intensity and source-to-detector distance; K_2 is the spherical cavity radius; and K_3 is the background response of the gage. These parameters are described in detail elsewhere (9).

The important assumptions made in arriving at this model are that (a) the gage geometry consists of a point source of monoenergetic fast neutrons surrounded by a spherical cavity of radius K_2 that is, in turn, surrounded by an infinite, homogeneous sample; and (b) the neutrons emitted by the source diffuse like gas molecules until they are removed from each energy group. It is not likely that this model describes very accurately the behavior of surface-type neutron gages except over very limited ranges of gage design parameters and sample compositions and densities. The gage geometry chosen is very artificial when applied to surface-type gages.

A more rigorous approach is possible with the depth-type neutron moisture content gages. Olgaard and Haahr (10) devised and tested a three-group neutron diffusion model for the depth-type gages that appears to be quite accurate for the gages and samples tested to date. A Monte Carlo model is presently being developed for surface-type gages in NCHRP Project 10-5A. This approach should prove to be quite accurate, but generally requires a large amount of digital computer time for each model prediction. It is possible that a simple model can be devised and tested based on the results of the Monte Carlo model.

THE CALIBRATION MODEL METHOD

To minimize the identified sources of error to acceptable levels, the calibration model approach was formulated in the initial phase of work on Project 10-5 (2). This approach consists essentially of developing a simple mathematical or calibration model for both nuclear gages that includes all the soil parameters that affect the gage response. This model contains constants that must be determined for each gage by a least-squares analysis of gage responses taken on samples with known characteristics. A set of calibration samples consisting of pure materials, such as aluminum and magnesium of known density and composition, is used with these models. The advantages of using such samples for calibration are that they can be chosen to be stable, homogeneous, and representative of typical soils.

The calibration model method is no longer required for the gamma-ray density gages because the dual-gage principle described later will be able to minimize satisfactorily the identified sources of error for those gages. However, a similar principle probably does not exist for the neutron moisture content gages, so that the calibration model approach should prove important in minimizing the sources of error for those gages. Therefore, the rest of the discussion of the calibration model approach is confined to the neutron moisture content gages.

The least sophisticated use of this method would be to calculate one calibration curve with the calibration model appropriate to one average soil composition and density. Even this relatively simple method of use should represent a considerable improvement over other previous methods of calibration. A slightly more sophisticated method would be to use the calibration model to calculate calibration curves representative of various densities of several soil types that can be visually identified, such as sand and clay. This method requires a knowledge of the soil composition as a function of the soil classification, and a knowledge of the sample density that can be obtained from a measurement with a gamma-ray density gage. This second method should be quite accurate, but it depends on the gage user's ability to identify visually the soil type and also on each soil type having a relatively constant composition. The most sophisticated use of the method would be to obtain the composition of the soil sample of interest and calculate a calibration curve from the calibration model specifically for that soil at various densities. This method would only be practical if the same soil is to be encountered for an extended period of time. This method still requires a measured value of the sample density. One shortcut method that would alleviate most of the work involved in compensating the gages for variable composition and density would be first to obtain the gage response to a soil sample of known density. Then the entire calibration curve for that soil would be back-calculated from the single point and the calibration model for the gage. This technique has not been tried yet, but should prove valuable.

Gamma-Ray Density Calibration Models

Although the calibration model approach is no longer necessary for the gamma-ray density gages, the calibration model developed for this type of gage may prove useful in the orderly optimum design and use of the dual-gage techniques that will probably replace the calibration model approach. Therefore, a discussion of the model that was developed in the second phase of work on Project 10-5 (9) and other possible models is given here.

The calibration model developed in Project 10-5 is given by

$$R = C \exp_{10} (a + bC + cP) \quad (4)$$

where R is the gage response; C is the Compton scattering probability; P is the photoelectric absorption probability; and a, b, and c are constants for a given gage that are determined by a least-squares analysis of gage responses taken on samples of known density and composition.

The Compton scattering probability is taken as

$$C = \rho \sum_{i=1}^{n} \frac{w_i Z_i}{A_i} \quad (5)$$

where ρ is the sample density; w_i is the weight fraction of element i; Z_i is the atomic number of element i; A_i is the atomic weight of element i; and n is the total number of elements in the sample. The photoelectric absorption probability is taken as

$$P = \rho \sum_{i=1}^{n} \frac{w_i Z_i^5}{A_i} \quad (6)$$

This sample model inherently assumes that one average gamma-ray path can be established for a given gage that is essentially constant over the range of sample compositions and densities that are to be encountered. In spite of the simplicity of the model, it has been found that it is quite accurate for a wide range of gage designs and sample compositions and densities. It has the additional advantage of being able to fit the boundary conditions $R = 0$, $\rho = 0$, and $R = 0$, $\rho \rightarrow \infty$. It has the disadvantage that it does not explicitly give the role of the gage design parameters, such as source-to-detector distance, source energy, collimation angles, and the detector efficiency, as a function of gamma-ray energy. These parameters are implicitly contained within the a, b, and c constants.

Prior to this study, very similar models to that of Eq. 4 were proposed by Irick (11) and Semmler et al (12), but they did not separate the Compton scattering probability from the photoelectric absorption probability. A recent paper by Czubek (13) describes a model similar to that given by Eq. 4 that attempts to extract explicitly the gage design parameters from the a, b, and c constants. This model may prove quite valuable in the orderly optimum design of dual-gage techniques.

The simple calibration model of Eq. 4 has proved to be quite valuable in leading to the discovery of the dual-gage principle and in optimizing the air-gap dual-gage method developed by Kühn (14). Gardner et al (15) applied the calibration model to air-gap responses and showed that the model gave a basis for a system of using the air-gap method that gave improved accuracy. A simple nomograph method of use was developed based on the proper application of the calibration model to air-gap responses. The details of this treatment and a method for determining the necessary nomograph are given by Gardner and Roberts (9) and by Gardner et al (15).

Neutron Moisture Content Gage Calibration Models

The neutron moisture content gage calibration model developed and used in NCHRP Project 10-5 is that given as Eq. 3, which also served as the detailed mathematical analysis model. For use as a calibration model, the K_1, K_2, and K_3 parameters in this model are determined for a particular gage by a trial-and-error analysis of gage responses taken on prepared samples. This model does not offer sufficient accuracy for use as a detailed mathematical analysis model, and has too much freedom of gage response shape to be used without additional information as a calibration model. The model might be useful as a calibration model if the gage response shape for a particular type of gage is established, and the factor K_2 in the model is restricted to values that will give rise to the correct gage response shape. It has been found that gage responses vs moisture content shapes can be concave, straight, or convex, depending on the value of K_2 that is used. Additional experimental work or results from the Monte Carlo model studies will be used to establish the gage response shape for particular values of the gage design parameters such as source-to-detector distance and the amount of moderator surrounding the source and detector.

DUAL-GAGE PRINCIPLE FOR GAMMA-RAY SCATTER GAGES

The dual-gage principle was discovered when gage responses to laboratory calibration samples from several different gages were being fitted to the calibration model given by Eq. 4. It became obvious that different gages had different relative sensitivities to the Compton scattering and photoelectric absorption probabilities. Because the primary effect of variations in soil composition is manifested in the photoelectric absorption probability, the possibility existed to use two gages simultaneously to determine density while eliminating the effect of soil composition by eliminating the photoelectric absorption probability. This is accomplished by obtaining the specific calibration model given by Eq. 4 for each of two different gages. If the calibration models are denoted for each of the two gages by the subscripts 1 and 2, then one obtains

$$R_1 = C \exp_{10} (a_1 + b_1C + c_1P) \tag{7}$$

and

$$R_2 = C \exp_{10} (a_2 + b_2C + c_2P) \tag{8}$$

From the definition of C, the density ρ can be extracted from these equations if it is assumed that

$$\sum_{i=1}^{n} \frac{w_i Z_i}{A_i} = 0.05 \tag{9}$$

Unfortunately, the simultaneous solution of Eqs. 7 and 8 is not straightforward, but several techniques are given by Gardner et al (15). A quadratic solution can be obtained if two terms of a series expansion of log ρ are used:

$$\rho = \frac{-B - (B^2 - 4AC)^{1/2}}{2A} \tag{10}$$

$$A = 0.05 (c_1b_2 - c_2b_1) \tag{11}$$

$$B = c_2 \log R_1 - c_1 \log R_2 - 1.6815(c_2 - c_1) + c_1c_2 - c_2a_1 + 6.5(c_1b_2 - c_2b_1) \tag{12}$$

$$C = 130c_2 \log R_1 - 130c_1 \log R_2 + 7.2384(c_2 - c_1) + 130(c_1a_2 - c_2a_1) \tag{13}$$

The density ρ in Eq. 10 is given in pounds per cubic foot. A solution can also be obtained from a nomograph and this procedure is outlined by Gardner and Roberts (9).

The advantage of the dual-gage principle is that it shows promise of being able to eliminate the effect of composition while not accentuating the effect of surface heterogeneities. It also has the capability of being implemented with existing gages or more efficiently with gages designed specifically and optimally for the dual-gage principle.

There are many possible methods for obtaining practical dual-gage systems. Any combination of two different source energies, source collimations, source-to-detector separations, detector efficiencies including the use of energy filters, and gage positions above the sample are possible. Some of these combinations can be used with existing gages and some can only be incorporated in new gage designs. Both types of dual-gage systems are discussed in the following. The optimum dual-gage technique is the one that will minimize the total error composed of those resulting from composition sensitivity, sensitivity to surface heterogeneities, and normal source emission fluctuations.

Use of Existing Gages

It is important that the dual-gage principle be capable of use with existing gages so that they can be used in an optimum fashion until a new generation reaches the market. Of the possible dual-gage techniques capable of use with existing gages, the air-gap method introduced by Kühn (14) is most promising. This technique consists of taking a gage response in the usual manner and then raising the gage to a fixed height above the sample surface where a second response is taken. A nomograph can be obtained that gives density independent of the sample composition as a function of the normal flush response and gap response. This technique is described in detail by Gardner and Roberts (9) and a sample application is described by Gardner et al (15).

To implement the air-gap method with the existing commercial gages, one only needs a jig for raising the gage to a predetermined height above the sample surface. Several of the gage manufacturers are now supplying such a jig with the gages. The gages can be easily calibrated with four laboratory samples. Suitable sample materials for calibration are described by Gardner and Roberts (9).

Other possible techniques are the use of two separate gages that inherently have different characteristics, and the use of energy filters, such as the placement of a thin lead sheet under an existing gage by a shutter mechanism of some sort. It is possible that one of these techniques would minimize the total error discussed in the previous subsection better than the air-gap method if it is found to be less sensitive to surface heterogeneities. This possibility is being studied in NCHRP Project 10-5A. To date only the feasibility of these other dual-gage techniques has been established.

Design of New Gages

The next generation of gamma-ray density gages will probably include optimally designed dual-gage systems. A major portion of the work in NCHRP Project 10-5A now under way is devoted to determining optimally designed dual-gage systems that will give minimum total error. The gage design parameters presently being studied for dual-gage implementation include source energy, source-to-detector separation, detector efficiency spectra, and source and detector collimation.

FUTURE WORK

Work is presently in progress on NCHRP Project 10-5A on several aspects of improving the design and use of nuclear gages. These include the optimum design of a dual-gage gamma-ray density system, a study of the energy discrimination technique for minimizing the composition effect of gamma-ray density gages, a study of possible methods of minimizing the surface heterogeneity effect on gamma-ray density gages, and the improvement of the analysis and calibration model for the neutron moisture content gages.

Two related programs of interest to nuclear gage users have recently been initiated. The International Atomic Energy Authority recently sponsored the writing of a guidebook on neutron moisture gages that should be published very soon. This guidebook will describe how the gages work, what the advantages and disadvantages of the gages are, state-of-the-art development of the gages, the role and present status of theortical analyses of the gages, and suggested methods of calibration and use of the gages. The other program, also endorsed by the International Atomic Energy Authority, is an extensive evaluation of the nuclear gages. This program is being carried out at Brno, Czechoslovakia, by the Czechoslovak National Association of the International Union of Testing and Research Laboratories for Materials and Structures. It consists of a series of tests to be performed on all commercially available nuclear gages including stability, temperature dependence, effective sample volume, and composition sensitivity. A report of the results of these tests will be published.

REFERENCES

1. Belcher, D. J., Cuykendall, T. R., and Sack, H. S. The Measurement of Soil Moisture and Density by Neutron and Gamma-Ray Scattering. U.S. Civil Aero. Admin. Tech. Dev. Rept. No. 127, 1950.
2. Ballard, L. F., and Gardner, R. P. Density and Moisture Content Measurements by Nuclear Methods—Interim Report. NCHRP Rept. 14, 1965.
3. Hughes, C. S., and Anday, M. C. Correlation and Conference of Portable Nuclear Density and Moisture Systems. Highway Research Record 177, pp. 239-279, 1967.
4. Goldstein, H., and Wilkins, J. E., Jr. Calculation of the Penetration of Gamma Rays. Final Report, U.S. Atomic Energy Commission, NYO-3075, June 30, 1954.
5. Taylor, D., and Kansara, M. Measuring Density With the Nuclear Back-Scatter Method. Nucleonics, Vol. 24, p. 54, 1966.
6. Taylor, D., and Kansara, M. A Theory of the Nuclear Densitometer. Soil Science, Vol. 104, p. 25, 1967.
7. Taylor, D., and Kansara, M. Improvements in the Method of Measuring the Density of Soil Using Nuclear Back-Scatter. Nuclear Instruments and Methods, Vol. 59, p. 305, 1968.
8. Semmler, R. A. Neutron-Moderation Moisture Meters. Final Report, U.S. Atomic Energy Commission, COO-712-73, Sept. 1963.
9. Gardner, R. P., and Roberts, K. F. Density and Moisture Content Measurements by Nuclear Methods—Final Report. NCHRP Rept. 43, 1967.
10. Olgaard, P. L., and Haahr, V. Comparative Experimental Investigations of the DM Neutron Moisture Probe. Nuclear Engineering and Design, Vol. 5, p. 311, 1967.
11. Irick, P. Applications of General Curve Fitting Procedure to AASHO Road Test Data. Highway Research Record 42, pp. 1-12, 1963.
12. Semmler, R. A., Bugger, J. E., and Rieke, F. F. Gamma Scattering Density Meters: Analysis and Design With Applications to Coal and Soil. Univ. of Chicago, Lab. for Applied Sciences, LAS-TR-161-35, Dec. 1961.
13. Czubek, J. A. Radioisotope Instruments in Industry and Geophysics. International Atomic Energy Authority, Vol. 2, Vienna, 1966.
14. Kühn, S. H. Effects of Type of Material on Nuclear Density Measurements. Highway Research Record 66, pp. 1-14, 1965.
15. Gardner, R. P., Roberts, K. F., Hughes, C. S., and Anday, M. C. A Calibration Model for Optimizing the Air-Gap Method of Compensating Nuclear Density Gages for Soil Composition Variations. Jour. of Materials, Vol. 2, No. 1, p. 3, 1967.

Embankment Compaction Variability—Control Techniques and Statistical Implications

T. G. WILLIAMSON, Research and Training Center, Indiana State Highway Commission

The development of a more effective method for field control of embankment compaction must be based on knowledge of the results being achieved using current inspection procedures. The purpose of this study was to determine the extent of compaction variability present in fill construction for typical Indiana construction projects and to identify the various factors that lead to this variation. Different techniques for measuring in-place density, including the sand cone replacement method, the water-filled balloon volume-measuring device, and the surface backscatter nuclear gage, were studied to prove variance estimates for use in the final statistical analysis.

The results indicate that widespread compaction variability is present in all field construction regardless of testing method, and that it is caused by a combination of many interrelated factors. This observed spread in compaction results indicates that current control procedures do not account for variability, and therefore an inspection program using statistical quality control procedures developed for these data is presented. The proposed technique is that of using a hypothesis decision theory that accounts for the compaction variability by using statistical parameters based on random sampling to make decisions as to overall compaction quality.

•THE PRESENT-DAY construction of highways is a high-speed, extremely complex operation. This is true whether it is a secondary road or a section of Interstate Highway. Because of this speed and complexity of construction, it is very important that adequate field control of the work be continually maintained to ensure that a quality product will be the end result.

One of the many important areas of construction over which this control must be maintained is the compaction of highway fills or embankments. For a fill to function as an adequate foundation for the structure it is to support, whether pavement or a bridge, the compaction must be controlled to ensure that the fill possesses given strength and stability characteristics and that these are uniform from one location to another.

To ensure that the finished fill is uniformly compacted to a specified level, it is necessary to establish a sampling and testing program that is related to the construction process and on which a realistic decision can be made with respect to the overall quality of the work. Historically, this control has been achieved by performing one or two control tests for a large quantity of material and accepting the entire volume of work as being satisfactory if the results of these isolated tests exceed some minimum specified value. The problem with this approach is that in most instances the sampling and testing program is not related to the true variability associated with the construction

Paper sponsored by Committee on Quality Assurance and Acceptance Procedures and presented at the 48th Annual Meeting.

process and the decision thus made on the basis of these control tests is at best an educated guess of the actual quality of the work.

A solution to this problem encouraged by the Bureau of Public Roads (2) is that of applying statistical quality control. However, a great deal of research data is required to provide the necessary information with regard to what variability exists using present-day construction methods if a statistical quality control technique is to be developed for highway construction. This study is an attempt to provide such data.

PURPOSE

Based on this need for research data related to the variability associated with current construction practices, the primary purpose of this study was to measure the variability found in the compaction of soils used in highway embankments. The ultimate goal was then to use this information to establish future sampling and testing programs that allow for these inherent variations.

One of the factors that contributes significantly to the overall observed variability is the field testing method employed. This is true whether the measurement being made is for field density, field moisture content, or the maximum standard density, to which the field density is compared in establishing a relative compaction level. To establish estimates of variability for these important soil characteristics, several different testing methods were employed to measure each of these parameters to provide comparative data between test methods. The methods used in this study were either procedures that are currently in use or are proposed as possible future test methods for Indiana State Highway control testing.

SAMPLING AND TESTING PROGRAM

Sampling Procedures

The general field sampling and testing procedures used in this study were based on guidelines established by the Bureau of Public Roads and described in the publication "The Statistical Approach to Quality Control in Highway Construction" (3). Three projects were selected that contained over 400,000 cu yd of fill construction and each of these was divided into a series of 10 individual fills containing a minimum of 5,000 cu yd. A series of replicate tests including in-place density, in-place moisture, and laboratory standard maximum density were performed in each control fill. Based on previous work done by the Indiana State Highway Commission in this area of compaction control (10), it was decided that a minimum of seven replicate field tests should be performed per control section to establish variance estimates, and that more tests than this would be desirable if field conditions permitted.

With respect to the actual test locations, a table of random numbers was used to establish a longitudinal station and lateral offset for each field test. The vertical location of the control test was obtained by instructing the field crews to perform the assigned tests in any compacted lift that had been passed by the project grade inspector as meeting specifications and that fitted their overall testing schedule with the restriction that only one test per lift be allowed. Thus, the field crews could test the control fills as they were constructed and certified as meeting specifications, thereby allowing them freedom to test in the areas where work was progressing.

Large 25-lb bag samples and smaller 200-gram samples were taken from each test location and used in a laboratory testing sequence to establish maximum density values and classification data for the materials encountered on each project.

Testing Procedures

The actual testing program included performing the tests of importance—that is, in-place density, in-place moisture content, and standard maximum density-moisture content—using a series of test procedures currently applicable to the control of field compaction.

The in-place density tests used during this study were the sand cone replacement method, the water-filled rubber-balloon volume-measuring device, and the nuclear density surface backscatter approach. The sand cone method used followed AASHO T 191-64 test procedures, and the balloon method followed AASHO T 205-64 specifications except for the fact that a pressure gage was not used to ensure constant water pressure for all tests. Instead of using a pressure gage, the technicians would pump the pressure in the cylinder to a point where no further change in water level occurred and then would take their readings.

Three different models of nuclear gages were employed, with each being assigned to a specific project during the course of the testing program. All gages used are of the backscatter design and are commercially available units. Details of the actual calibration of these gages are presented in a paper by Williamson and Witczak (11).

The field testing procedure involved first performing the necessary nondestructive nuclear counts using both the density and moisture probes on the selected test locations. The technicians then augered a hole for the balloon measurement in the exact location where the nuclear readings were taken and determined the density in this manner. Because the balloon method left the original hole intact, it was next possible to perform the sand cone test using the same hole or by augering it out to a slightly larger diameter to fit the sand cone plate being used. Both of these approaches were employed.

This testing procedure permitted a comparison to be made between the three methods on essentially the same material, although the influence of the nuclear gage was effective over a larger volume than the other two methods.

After the soil had been removed from each test hole, moisture determinations were made using both a Speedy carbide-gas moisture tester and the conventional laboratory oven-drying technique. These were in addition to the previous moisture determination as obtained by the nuclear equipment.

The portion of the sample from the density hole remaining after the moisture determinations had been made was used to perform a field one-point maximum density compaction test. This method is used by many agencies, and previous research by the Indiana State Highway Commission (10) has indicated the relative merit of such a test in establishing maximum density values for soils in the field. A unique feature of this test is that only one point of the standard AASHO T 99(A) compaction curve is established and this is accomplished under field conditions. The density and moisture content of this individual compaction test point are plotted on a set of typical compaction curves developed for Indiana soils and the maximum dry density and optimum moisture content are thus defined.

Because the one-point compaction test has not been approved for routine field use in Indiana, samples were taken from the vicinity of the replicate sand cone holes and subjected to a complete laboratory compaction test according to AASHO T 99(A). This provided data to compare these two approaches for determining maximum density and optimum moisture content.

Projects Tested

The three projects selected for this study were chosen to provide a wide degree of variation with respect to soil homogeneity and compaction technique used. All three projects were characterized as Interstate high-type rigid pavement construction. All of the projects are located in the glaciated till plains section of the Central Lowlands Province as defined by Lobeck (8).

Project 1 is located in central Indiana in an area of little topographic relief with the soil being geologically classified as a Tazewell stage, Wisconsin age glacial drift. The soils tested on this project were relatively homogeneous, consisting primarily of a low-plasticity silt classified by the HRB procedure as an A-4(5) with some isolated silty clays, A-6(7), also present. Compaction was achieved by the use of a towed sheepsfoot roller.

Project 2 is located in west central Indiana in an area of relatively dissected rolling topography. The soils are characterized as Illinoian age glacial till and are relatively heterogeneous, ranging from a low-plasticity silty clay, A-4(6), to a moderately plastic clay, A-6(9). Compaction on this project was achieved using a combination of a self-propelled sheepsfoot and a rubber-tired roller.

Project 3 is located in a very level area of northwestern Indiana in the Cary stage, Wisconsin age glacial drift region. This is a very young drift with the soil characterized as very heterogeneous, ranging from highly plastic lacustrian deposits, A-7-6(12), to granular beach sand deposits, A-2-4(0). Compaction equipment included towed sheepsfoot units, rubber-tired rollers, and steel wheel rollers, all used concurrently in a given control section.

CORRELATION TESTING RESULTS

Determination of Standard Maximum Density

The current Indiana State Highway Commission practice for establishing maximum density values for a given construction project is to obtain a series of representative soil samples from the project, and then to determine maximum density according to the standard AASHO T 99(A) compaction test at a central laboratory. The results of these tests are then sent to the project personnel and it is their responsibility to apply the correct values to the various field conditions. Because the soil varies widely from one location to another, this becomes an extremely difficult task for the field inspector. The results of the laboratory compaction tests for the samples collected during this study indicated that a range in maximum density values of about 20 pcf (pounds per cubic foot) existed for all three projects, thus emphasizing the importance of being able to determine the correct maximum density value to be used in computing relative compaction values for a given in-place field density test.

To avoid this problem of selecting a control density value, a field one-point compaction test as previously mentioned was performed for each field test location. The results of these field tests were used to determine the maximum density and optimum moisture content for each sample by comparing the observed density and moisture values with a set of typical Indiana curves as developed by Walter T. Spencer, Chief of the Indiana State Highway Commission Division of Materials and Tests.

The field one-point compaction test has several advantages over the present Indiana method of representative field sampling and laboratory testing. The test can be performed on the grade in about 10-15 minutes and requires only a minimum amount of additional equipment, thus resulting in a savings in time and money when compared to the prospect of extensive field sampling and laboratory testing. An additional advantage is that the one-point test establishes maximum density for the material from the in-place density test location itself and the density thus obtained can be used with more assurance that it is representative of the material.

A comparison of the laboratory and field one-point values obtained for standard maximum density indicates the one-point values average 2.5 pcf less than the laboratory data, based on 436 observations. A similar value of 3.1 pcf was obtained during a previous study (10) of subgrade compaction variability. Probably the major reason for this deviation is the fact that during the laboratory compaction test the sample is re-used for each subsequent test point on the compaction curve, whereas the field test involves the use of a new sample for each test point.

The discussion indicates the relative merit of the field one-point compaction test and the data obtained for this test compare favorably with the laboratory test results. On this basis, most percent compaction data presented in this study are based on these field one-point maximum density test values.

Measurement of In-Place Wet Density

The current method preferred by the Indiana State Highway Commission for measuring in-place density is the sand cone. This method has been in use for many years and is a proven field technique. However, there has recently been some dissatisfaction with this method because of (a) the amount of time required to perform the entire test, (b) the possibility of making an error at any one of the many steps associated with the test, and (c) the necessity of making a series of detailed computations to arrive at the final density value. Based on this, two alternate approaches, the balloon volume-

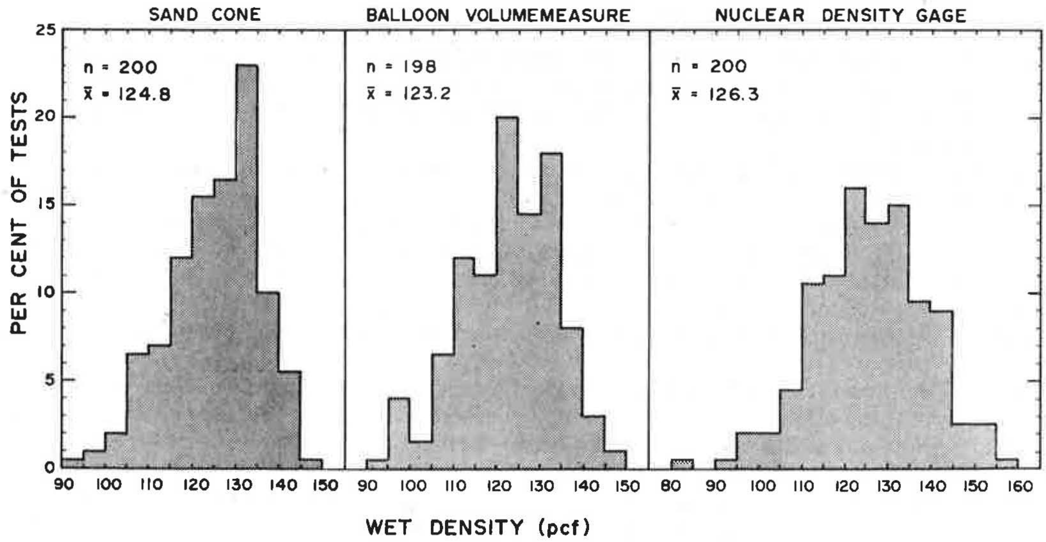

Figure 1. Variation of in-place wet density values for Project 1.

measuring method and the nuclear backscatter method, were also employed to measure field density.

The balloon device is used by many agencies and the use of nuclear equipment is rapidly gaining favor among highway engineers. Both have the advantages of simplicity with respect to computations involved and relative speed of operation in comparison with the sand cone technique.

The subject of developing appropriate calibration curves for nuclear moisture-density gages has been dealt with by many authors, with several different methods proposed (5, 11). The calibration technique used by the Indiana State Highway Commission is based on the work done by Williamson and Witczak (11) and basically involves using a statistical hypothesis testing approach. Three different brands of nuclear gages were used of which only one performed satisfactorily. A series of electronic failures that were not easily corrected made it impossible to collect sufficient field data for the nuclear density units assigned to Projects 2 and 3.

The results of performing all three tests on Project 1 are shown in Figure 1, which shows the variations in field wet density values. These data indicate that the methods all resulted in similar values of density, as shown by the average values ranging from 123.2 pcf for the balloon device to 126.3 pcf for the nuclear gage. Also, the overall range in values, from 90 to 150 pcf, was approximately the same for all three techniques.

Comparative results for the balloon and sand cone data obtained on Projects 2 and 3 indicated similar results, although for Project 2 the balloon values were approximately 5 pcf lower than the sand cone values, whereas the average values for Project 3 were almost identical for the two methods. A possible explanation is that, on Project 3, the field operators used exactly the same density hole for both sand cone and balloon tests, whereas on the other two projects the technicians enlarged the density hole to fit the sand cone apparatus after taking their balloon test. The enlarged sand cone test hole was approximately 0.05 cu ft, compared to about 0.03 cu ft for the balloon density test.

The overall close similarity of results obtained for these three field density tests indicates that any of them could be used in the field with equal reliability. The decision as to which method to use would then depend on which particular advantage or disadvantage associated with the tests was felt to be most critical or important. The nuclear gage would be best suited for performing an extensive series of tests, as might be required for a statistical quality control program, because of its speed of operation.

Determination of Moisture Content

The computation of dry density, which is used as the basis for computing percent compaction in this study, depends on the accurate determination of the moisture content of the wet density sample. This is true for both the in-place field and standard laboratory density tests.

The two common methods presently used for field measurement of moisture content in Indiana are the field stove-drying method and the Speedy carbide-gas moisture tester. The stove-drying method has been in use for many years and a previous study conducted by the Indiana State Highway Commission (10) indicated a high degree of correlation with standard laboratory oven-drying results. The Speedy moisture test is relatively new and no significant amount of documented correlation data for its results had been obtained by Indiana State Highway Commission field personnel prior to this study.

A third method for determining field moisture content being used in Indiana on a trial basis employs the nuclear surface backscatter moisture gage. Based on the need for data to determine the applicability of these methods, moisture data obtained using both units were compared with standard laboratory oven-drying results. The desire to use either of these methods for field control is based on the speed at which the test can be performed in comparison with oven-drying or field stove-dryings methods, and the simplicity of the computations associated with both methods.

The results obtained during this study indicate that both methods provide results that compare very closely to laboratory oven-drying data. A linear correlation analysis was performed for the data, and correlation coefficients from 0.81 to 0.85 were obtained for the nuclear gages, indicating the validity of using this type of equipment. It should be pointed out that the nuclear gage is affected by the total hydrogen (moisture) in the soil, whereas the laboratory oven-drying method accounts for only the moisture that can be driven out of the sample at a temperature of 105 C. Also, the nuclear gage measures the average moisture content of a relatively large volume of material, whereas the oven-dry samples represent only 150 to 200 grams of soil. These differences are assumed to account for the major portion of the observed variations when comparing moisture contents determined by these methods.

Results obtained comparing moisture content determined by the Speedy device with oven-drying results also indicated excellent correlations. The data for this phase of the study are based on obtaining moisture samples from two different test sources, the actual in-place density test material and the material extracted from the one-point compaction test sample. Also, two different sizes of Speedy moisture units, one with a 6-gram sample capacity and the other using a 26-gram sample, were tested. One of the major criticisms of the Speedy method is that the relatively small size of sample used does not provide representative results and this use of two sample sizes allowed a comparison to be made with respect to the results achieved by each unit.

The data indicate that very high linear regression correlation coefficients were achieved when using the 26-gram sample size Speedy moisture tester for all field test data. These ranged from 0.84 to 0.93 and appeared to be independent of sample sources, i.e., in-place material or one-point compaction test sample. The results for the smaller 6-gram sample tests were not as encouraging, especially for the in-place density material, which had a correlation coefficient of only 0.48 compared with 0.88 for the one-point compaction test material. A possible explanation lies in the relatively uniform moisture content present in a one-point compaction test sample as prepared by the grade inspector vs the possible nonuniformity of moisture content that may be present in the in-place density material.

Thus, based on the preceding correlations between the field methods tested and laboratory oven-drying values, it is suggested that either the larger (26-gram) Speedy moisture tester or the nuclear technique can be used to establish reliably estimates of moisture content. Both methods can be performed in a very short period of time, so that either would be applicable to a statistical quality control testing program that places an emphasis on speed because of the number of tests required.

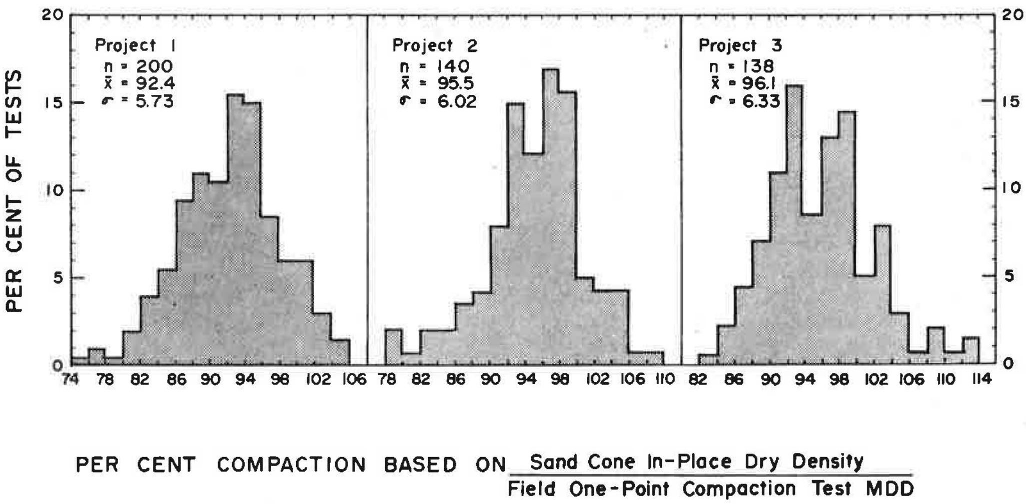

Figure 2. Variation of field compaction for study projects.

PERCENT COMPACTION RESULTS

Current Indiana State Highway Commission specifications (6) require that all fill material be compacted to a minimum of 95 percent of AASHO T 99(A) maximum density. This latter density value is obtained by testing representative field samples obtained for typical project soils. Because this approach permits some area of doubt with respect to whether or not the soil from the field test hole actually exhibits the compaction characteristics of these representative samples, the alternate method of using the field one-point compaction test to determine maximum density was used for determining percent compaction levels for this study.

Using the dry density data obtained from the sand cone test, which is the method currently used by field personnel, and the corresponding one-point maximum dry density values, the distributions of relative compaction for the projects studied are shown in Figure 2. The first values of interest are those for mean relative compaction. These are observed to be 92.4, 95.5, and 96.1 for Projects 1, 2, and 3, respectively. These indicate that for Projects 2 and 3, the contractor met specifications based on the average of the tests, whereas the average compaction level of Project 1 was well below minimum specifications.

The second values of interest are the overall ranges in compaction that were observed. In general, the data for all projects showed a similar trend in variability with low values of from 74 to 82 percent compaction up to maximum values of 106 to 114 percent compaction depending on the test method involved. This indicates a general range of almost 30 percent relative compaction values for all three projects, illustrating the extreme variations that must be accounted for in a realistic control program.

A third value that expresses a characteristic of the distribution of the data is the standard deviation. The standard deviation is a statistical expression that provides an indication of the variation of the data about the overall mean of the distribution. This is to say that if the data points are close to the average, the standard deviation will be small in magnitude. The standard deviations indicated by Figure 2 are 5.73, 6.02, and 6.33. These values indicate that, while the mean compaction level for Project 1 was lower than for the other two, the compaction was slightly more uniform. Because it is a combination of two parameters—actual relative compaction level and uniformity of compaction—that determines whether a volume of soil has been compacted satisfactorily, it is difficult to establish which of the three combinations observed in this study actually represented the best overall compaction.

TABLE 1

SUMMARY OF PERCENT COMPACTION RESULTS FOR STUDY PRODUCTS

$\left(\text{Percent Compaction Based on } \frac{\text{Indicated Test In-Place Dry Density}}{\text{Field One-Point Compaction Test MDD}}\right)$

Category	Project 1			Project 2		Project 3	
	Sand Cone	Balloon	Nuclear	Sand Cone	Balloon	Sand Cone	Balloon
Number of sampling locations	100	99	99	70	70	69	67
Number of compaction determinations	200	197	198	140	140	138	134
Range of percent compaction data	74 to 106	70 to 108	74 to 118	78 to 110	70 to 108	76 to 116	78 to 116
Average percent compaction	92.40	90.80	93.48	95.46	90.92	96.05	96.80
Standard deviation	5.73	6.63	7.48	6.02	7.25	6.33	6.13
Percent of tests less than specification limit of 95 percent compaction	67.0	74.5	57.5	43.5	70.7	50.0	46.2

The fourth important characteristic of these distributions is that they are all normal. The Kolmogorov-Smirnov test (9) for normality was applied to the three projects and they were all found to be normally distributed at the 95 percent significance level. This was also true for all percent compaction data obtained during this study regardless of the test methods used.

It should be noted that the data for Projects 2 and 3 were obtained by performing the specified seven replicate density tests in each of the ten control sections. A larger number of data values was obtained for Project 1, because a favorable construction schedule made it possible to perform ten replicate tests in each control section. Thus, the distribution of data for this project represents a larger random sample from the total infinite population than for the other two projects. A summary of the overall compaction data obtained during this study is given in Table 1.

For Project 1 the range in percent compaction data is observed to be approximately the same for all three methods, varying from 75 to 108 percent, but the standard deviations of the data are quite different. These values range from 5.73 for the sand cone to 7.48 for the nuclear gage, with a value of 6.63 for the balloon tests. These data indicate that, although the results for the nuclear gage show the highest average level of compaction, they also have the largest degree of variability around their mean. It is noted that average compaction levels are below specifications for all three methods of measuring in-place density. The compaction results for Project 1 are also shown graphically in Figure 3 with the normal distribution curves superimposed on the frequency histograms.

The Project 2 mean compaction level of 95.5 percent for the sand cone data exceeded the specification limit and was significantly higher than the average compaction level of 90.9 percent recorded for the balloon data. Also, of the total number of tests performed, 56.5 percent of the results based on the sand cone test exceeded specifications, whereas this value was only 29.3 percent for the balloon test data. The standard deviation of 6.02 for the sand cone was much lower than the balloon value of 7.25, indicating the sand cone data resulted not only in a higher average level of compaction but also in a more uniform situation. However, because the true density at each test location was not established, it is impossible to say which of these sets of data is the most representative of the true compaction level.

The overall results obtained for Project 3 indicated close agreement between the sand cone and balloon data, with mean compaction values of 96.1 and 96.8 percent respectively. Also, the standard deviation values are almost identical, being 6.33 for the sand cone and 6.13 for the balloon data. The percent of total tests that exceeded the minimum specification limit was slightly over 50 percent for both methods.

It should be noted that for the field testing on this project, the sand in-place density test hole was used for both methods, indicating that either one will be satisfactory when applied to identical field conditions.

A summary of compaction results for the individual control sections of Project 1 is given in Table 2. The average compaction levels vary considerably from one section to another regardless of the in-place density test involved. A similar variation is also observed for the standard deviation values, with the sand cone data exhibiting the most consistent results for both average compaction and uniformity, the latter as given by the standard deviation estimates. Similar variations in compaction data were also obtained for the other two projects when comparing individual control sections.

ANALYSIS OF VARIANCE RESULTS

General Variance Terms

The basic mathematical technique known as the analysis of variance, denoted as ANOV, used in this study was a one-way Model II, equal number of observations per treatment approach. This ANOV was applied to relative compaction expressed as a ratio of field dry density to one-point compaction test maximum dry density.

It should be pointed out that to apply the ANOV technique, the data must first satisfy the criteria of being normally distributed with homogeneity of variances. The Kolmogorov-Smirnov test for goodness of fit (9) was used to test for normality, and results indicated that the percent compaction data for all projects was normally distributed at the 0.05 confidence level. The Foster-Burr test for homogeneity of variance (4) was applied to the ten control sections within each project and the results of this analysis indicated that the percent compaction variances were homogeneous at the 0.05 confidence level.

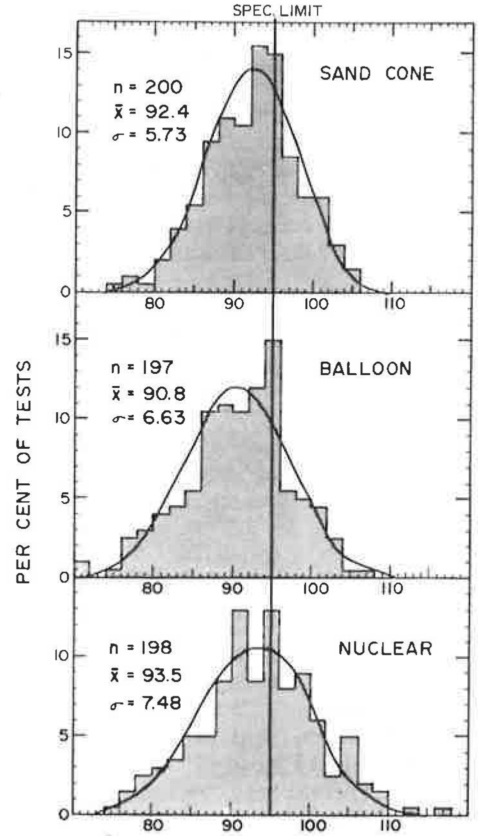

Figure 3. Variation of field compaction for Project 1.

TABLE 2

SUMMARY OF PERCENT COMPACTION DATA FOR PROJECT 1

(Percent Compaction Based on $\frac{\text{In-Place Dry Density}}{\text{Field One-Point Compaction Test MDD}}$)

Control Section	No. of Tests	Sand Cone Mean	Sand Cone Std. Dev.	No. of Tests	Balloon Mean	Balloon Std. Dev.	No. of Tests	Nuclear Mean	Nuclear Std. Dev.
1	20	89.91	7.45	18	85.92	8.12	20	92.33	7.84
2	20	94.02	6.76	20	91.67	8.53	20	90.65	6.78
3	20	92.64	4.47	19	92.71	4.66	20	95.10	7.57
4	20	90.79	5.85	20	88.92	7.72	20	96.83	6.67
5	20	92.05	5.76	20	89.50	5.13	20	91.56	5.87
6	20	93.78	4.82	20	92.78	6.43	20	94.61	8.87
7	20	93.07	5.89	20	91.72	6.17	20	92.11	8.19
8	20	93.85	5.73	20	92.77	6.12	20	95.20	5.08
9	20	91.17	5.43	20	90.60	5.28	18	95.40	7.00
10	20	92.69	3.97	20	90.36	4.98	20	91.22	8.78
OVERALL	200	92.40	5.73	197	90.80	6.63	198	93.48	7.48

The analysis of variance establishes values for three important terms. These are the within-treatment variance, the between-treatment variance, and the standard deviation estimate based on the combination of the preceding variances.

The within-treatment variance, denoted as σ_ω^2, represents the variation in compaction between replicate tests. The magnitude of this variability is basically a function of (a) "testing error" including the inherent inconsistencies in the field tests themselves and the error introduced by the individual technician performing the test, (b) material variations within a relatively small testing area, and (c) variability associated with the compaction process.

The between-treatment variance denoted by σ_β^2 represents variations in compaction between treatments along the project with a treatment defined as a pair of replicate tests performed at essentially the same location. Again, as for the within-treatment variance, the factors primarily influencing the magnitude of the between-treatment variance are (a) soil type variations, (b) variations associated with the actual compaction process used, and (c) technician or testing variability, arranged in order of importance.

The standard deviation estimate accounts for both variability within small test areas and the variability between these test areas throughout the fill. This value is computed by using the following relationship for the overall variance estimate (σ_ϵ^2):

$$\sigma_\epsilon^2 = \sigma_\omega^2 + \sigma_\beta^2 \quad (1)$$

Overall Project Variances

Because each contractor was required to achieve the same minimum level of compaction for all of the fill earthwork, the ANOV was first performed on the total data for each project. Using the sand cone data for comparative purposes, the results show that the within-treatment variance term is relatively constant, ranging from 13.07 for Project 3 to 16.91 for Project 1. Because this term basically represents the testing error, it would be expected to be constant, assuming the field technicians all performed the tests by following carefully prescribed procedures.

The between-treatment variance terms vary widely from a minimum value of 15.97 for Project 1 to 21.81 for Project 2, and finally to a maximum of 27.24 for Project 3. These between-treatment variance results for the three projects emphasize the importance of the effects of soil variability and compaction process variations in analyzing compaction variability. This is evidenced by referring to the previous section describing project selection, which indicated the relative homogeneity of soil characteristics and construction techniques for Project 1 in comparison to the very heterogeneous nature of Project 3.

Variances Associated With Field Density Tests

The ANOV was also performed on the data obtained for each individual fill control section using results for each field density test method used. These results provide an indication of the compaction variability for individual fill areas while also providing a comparison of the various field testing techniques. An example of this variance data is given in Table 3.

Examining these results for Project 1, the within-treatment variance varied considerably from section to section and method to method. In particular, this variance term appeared to be very high for control section 1, especially for the balloon and sand cone compaction results. This is explained by the fact that the initial field tests performed by the personnel assigned to this project were all in this section because the contractor concentrated on this fill volume at the beginning of the testing phase. The fact that the technicians had not yet sufficiently developed their testing techniques obviously contributed significantly to the occurrence of these large within-treatment variances. As testing experience was gained, these variances decreased considerably, as shown by Table 3.

The within-treatment variances for the nuclear gage did not exhibit this pehnomenon, probably because of the minimal influence that the operator has on this type of density

TABLE 3

SUMMARY OF ANOV RESULTS FOR RELATIVE COMPACTION DATA OF PROJECT 1

$\left(\text{Percent Compaction Based on } \frac{\text{In-Place Dry Density}}{\text{Field One-Point Compaction Test MDD}}\right)$

Control Section	Sand Cone			Balloon			Nuclear		
	σ^2_ω	σ^2_β	σ_ϵ	σ^2_ω	σ^2_β	σ_ϵ	σ^2_ω	σ^2_β	σ_ϵ
1	56.96	–	7.56	64.33	3.28	8.20	62.76	–	7.90
2	25.17	21.72	6.85	35.24	39.62	8.62	40.88	5.38	6.79
3	11.69	8.78	4.51	15.70	26.14	6.46	15.37	44.31	7.71
4	16.63	18.64	5.93	29.36	32.40	7.84	28.12	17.26	6.72
5	7.21	27.46	5.89	20.68	5.96	5.16	11.19	24.50	5.96
6	16.05	7.61	4.86	13.19	29.78	6.53	61.80	17.86	8.90
7	6.02	30.25	6.02	23.16	16.20	6.26	20.85	48.75	8.32
8	12.24	26.76	6.24	14.83	23.92	6.21	9.07	17.67	5.16
9	6.96	23.75	5.54	11.24	17.57	5.36	43.20	6.18	7.01
10	10.13	5.96	4.02	17.00	8.27	5.02	48.08	30.71	8.85
OVERALL	16.91	15.97	5.71	24.15	19.67	6.60	34.04	21.99	7.48

test, but the variances were more widely scattered than for the sand cone and balloon data. A comparison of the within-treatment values obtained for different density tests for all test projects indicates that in general the lowest values are obtained for the sand cone.

The between-treatment variances for all three projects appear random in nature and vary widely from section to section within each project studied. The highest values were found for Project 3, which has been described previously as characterized by variable soil conditions and compaction techniques.

PRACTICAL APPLICATION OF A STATISTICAL QUALITY CONTROL PROGRAM

Basic Concepts

Most current highway specifications imply a form of quality control. A sampling and testing program is usually applied to the finished product, and a decision is made concerning the quality of the construction on the basis of these tests. Unfortunately, the number of compaction tests involved is usually only one or two and the results of these are taken as being representative of a relatively large volume of material. To be sure that the true compaction level of a given enbankment or fill had been established would require performing an infinite or at least an extremely large number of field tests and this would not be realistic. The use of a statistical control program is then a compromise between these two situations. Statistical control of a construction process involves using some statistical technique to make a decision about overall quality based on results of a random sample.

Before considering which statistical technique might be most applicable to embankment compaction control, a decision must be made with respect to the size of the control section in which the specified random sample tests are to be performed. A comparison of the variances obtained from the ANOV for each fill with the corresponding volume of the fill, which ranged from 7,500 to over 100,000 cu yd, indicated little or no correlation between these parameters. This agrees with previous data collected during a study of subgrade and subbase compaction (10).

Several approaches can be used in selecting the size of the control section. One method is to establish a fixed volume (or area) of material to be used as the control section, thereby ensuring equal control testing of all materials involved. However, it is difficult to estimate an optimum volume or area to be used as a control section based on the relative independence that seems to exist between observed variability and the corresponding testing area within which the variation is recorded.

An alternate approach is to base the size of the control section on the individual fills or construction units within the project, as was done in this study. Before construction begins, the different construction areas can be established by their stationing and each of these would then constitute a control section. However, it is suggested that for larger fills, say over 2,000 ft in length, the fills should be subdivided and these smaller units then used as the control sections. This approach generally coincides with the contractor's schedule in that he will usually construct a fill by placing and compacting a given lift over the length of the fill unless it is relatively long. If it is too long, it may be more feasible to build up several lifts over a portion of the fill and then proceed with the construction of the remainder of the fill.

There are many statistical techniques available that can be applied to this problem of relating a sample mean to the true mean of a normal population. One approach that can be practically applied to the control of compaction and that is recommended by this author is that of a hypothesis testing procedure based on a t statistic. An example of the use of this type of analysis applied to the fill compaction data from this study is presented in the following section.

Statistical Control Based on Hypothesis Testing

A statistical hypothesis is a statement about the value of some population parameter such as the mean or standard deviation. The parameter of interest in this study is that of the true mean value of relative compaction for a specified fill lift or population.

The hypothesis to be tested in the case of embankment compaction control is H_0 : $\mu \geq \mu_0$ where μ represents the true population mean and μ_0 is a specified acceptable compaction level. Because it is impossible to test the entire population, the decision to accept or reject the null hypothesis, H_0, must be based on the statistics of a randomly selected sample. Thus, based on the mean, $\bar{X}$, of a specified number of test samples, n, a decision is made relative to whether or not the true population mean, μ, from which these samples were randomly selected exceeds some specified value, μ_0.

Two values that must be established in any hypothesis test are the probability of making a Type I error, which is the rejection of the hypothesis when it is really true, and the probability of making a Type II error, or accepting a false hypothesis. The probabilities are denoted by α and β respectively, and a value of 0.05 was assumed for both of these parameters, thus compromising between the risk accepted by the highway department, β, and that accepted by the contractor, α.

Another critical value involved in hypothesis testing is the estimate of the population variance or standard deviation. Because it is impossible to establish a value for the true population variance, an estimate must be used. For this study project, standard deviation values ranged from a minimum of 5.7 for the sand cone data of Project 1 to a maximum of 7.5 for the nuclear data of that project. To account for the range in values obtained for the different test methods and projects studied, the following values are proposed for this type of construction: sand cone, σ_ϵ = 6.0; balloon, σ_ϵ = 6.7; nuclear, σ_ϵ = 7.5. These estimates represent three-dimensional variability rather than variability for a given lift, and comparative data obtained from a previous study (10) involving two-dimensional sampling indicated a relative compaction standard deviation estimate of approximately 5.0 as being realistic for sand cone testing.

The decision as to the value to be assigned for μ_0 or the specified compaction control level is subject to question. Current specifications require that all field compaction exceed 95 percent of standard AASHO maximum density. However, the results of this study, obtained by performing tests only in areas that had previously been passed as meeting the 95 percent compaction specification based on the field inspectors' tests, show that a considerable percentage of the tests fell below this level.

As previously noted, the relative compaction data were observed to be normally distributed, thus indicating that approximately 68 percent of the total data points are between the mean and plus or minus one standard deviation. Based on this concept, and using the means and corresponding standard deviation estimates for the three study projects, it can be shown that approximately 16 percent of all field data fell below 88 percent compaction. By changing the specification limit to 102 percent by increasing it by one standard deviation, this would result in 16 percent of the total data falling below only 95 percent compaction, which would be a more desirable situation. However, whether this is necessary or not is debatable because construction quality obtained by using the current 95 percent specification appears to be satisfactory.

Based on the indicated values of α, β, and σ_ϵ for each density test method, it is possible to determine the number of field tests that would be required for this approach based on a t test for the significance of means. This is accomplished using Appendix 9 of "Statistics in Research" (9). In using this table, a value of 7.0 was selected for δ, which represents the minimum level at which the hypothesis test is to be detected.

Based on these data, the number of field control tests required using the different field density testing methods is 10 with the sand cone, 11 with the balloon, and 14 with the nuclear gage used in this study.

TABLE 4

TYPICAL COMPUTATIONS FOR STATISTICAL DECISION THEORY USING PSEUDO t STATISTIC

Field Relative Compaction Data:

$X_1 = 104.2$	$X_6 = 97.7$
$X_2 = 96.6$	$X_7 = 98.6$
$X_3 = 89.4$	$X_8 = 94.9$
$X_4 = 92.3$	$X_9 = 90.8$
$X_5 = 93.0$	$X_{10} = 101.4$

Computations:

$H_0: \mu \geq \mu_0$ with $p = 0.95$, $n = 10$, $\mu_0 = 95.0$

$$X_{max} = 104.2$$

$$X_{min} = 89.4$$

$$R = 104.2 - 89.4 = 14.8$$

$$\tau = \frac{\frac{(104.2 + 89.4)}{2} - 95.0}{14.8} = \frac{96.8 - 95.0}{14.8}$$

$$\tau = \frac{1.8}{14.8} = 0.121$$

τ tabular $= -0.22$

$\tau > \tau$ tabular

Therefore, the embankment compaction does meet specifications.

The routine control procedure would be to perform the above number of tests for each lift and then compute a test statistic t using

$$t = \frac{(\bar{X} - \mu_0)\sqrt{n}}{s} \tag{2}$$

The decision would then be made to accept the hypothesis $\mu \geq \mu_0$ if the calculated t value equaled or exceeded a negative tabular t value as obtained from a cumulative t distribution table.

Unfortunately, using from 10 to 14 test points would make the computation of the mean and standard deviation relatively time-consuming, thus unnecessarily slowing construction. To simplify field computations, it is proposed that a pseudo t statistic denoted as τ be used. This value is computed by the relationship

$$\tau = \frac{\left(\frac{X_{max} + X_{min}}{2} - \mu_0\right)}{R} \tag{3}$$

where X_{min} and X_{max} represent the minimum and maximum relative compaction values for a random sample of n tests, and R is the range between these values. This calculated τ value is then compared to a tabular critical value as given by Appendix 17, Table 3 of "Statistics in Research" (9). The hypothesis is then accepted if τ calculated $\geq -\tau$ tabular. It must be noted that a random sample from a normal population is assumed. An example illustrating this approach is given in Table 4.

CONCLUSIONS

The results of this study and other similar quality control investigations clearly illustrate the importance of adopting a more realistic field sampling and tesing program if adequate control of compaction construction is to be gained. The sophisticated procedures now being applied to highway design coupled with the speed of current highway construction have antiquated the inspection control now being enforced.

The overall wide variations in relative compaction results observed in this study indicate the need for more extensive field testing coupled with a statistical technique for making a rational decision concerning overall compaction quality. This compaction variability is a combination of many interrelated factors including material variations, testing error, and compaction methods, and can only be accounted for by applying a sampling and testing program that accounts for this variability.

The approach suggested in this paper is the use of a hypothesis test using a pseudo t statistic for making the decision to accept or reject the work. This technique allows a rational decision to be made concerning the overall quality of compaction based on the results of a series of random tests rather than trying to judge the quality on the basis of one or two isolated test results.

REFERENCES

1. Standard Specifications for Highway Materials and Methods of Sampling and Testing: Part II, Method of Sampling and Testing. Ninth Edition. AASHO, 1966.
2. Quality Control and Acceptance Specifications. Proc., Highway Conference on Research and Development of Quality Control and Acceptance Specifications, Vol. 1. Bureau of Public Roads, April 1965.
3. The Statistical Approach to Quality Control in Highway Construction. Bureau of Public Roads, April 1965.
4. Foster, L. A. Testing for Equality of Variances. PhD thesis, Purdue Univ., 1964.
5. Hughes, C. S., and Anday, M. C. Correlation and Conference of Portable Nuclear Density and Moisture Systems. Highway Research Record 177, pp. 239-279, 1967.
6. Standard Specifications. Indiana State Highway Commission, 1963.
7. Joslin, J. G. Ohio's Typical Moisture-Density Curves. American Society for Testing and Materials, Reprint, STP No. 239.
8. Lobeck, A. K. Map of the Physiographic Provinces of North America. The Geographical Press, 1948.
9. Ostle, B. Statistics in Research. Iowa State Univ. Press, 1963.
10. Williamson, T. G., and Yoder, E. J. An Investigation of Compaction Variability for Selected Highway Projects in Indiana. ISHC Research Rept., Dec. 1967.
11. Williamson, T. G., and Witczak, M. W. Factors Influencing the Application of Nuclear Techniques to Soil Compaction Control. Highway Research Record 177, pp. 112-135, 1967.
12. Yoder, E. J., and Woods, K. B. Compaction and Strength Characteristics of Soil Aggregate Mixtures. HRB Proc., pp. 511-520, 1946.

Measuring the Variability of Compacted Embankments

JAMES L. JORGENSON, Engineering Experiment Station,
North Dakota State University, Fargo

This report contains the methods and results of measuring the variability of percent compaction and moisture content of compacted embankments in acceptable highway construction in North Dakota. Separate from the State Highway Department's control sampling, randomly located samples were taken on each of three typical construction projects. At each sample location, the following duplicate samples were taken: (a) in-place density using the water-balloon method; (b) in-place moisture by drying two soil samples; (c) moisture and density using a nuclear moisture-density gage in both direct transmission and backscatter positions; and (d) a sack sample for determination of maximum density. For comparison, State Highway Department data collected during the same construction period are also reported.

The major conclusions are that (a) the variability in percent compaction is large—for example, every third sample will deviate from the average by at least 3 to 5 percent; (b) the average percent compaction was very near the required minimum; (c) the higher in-place densities and lower standard deviations of the highway department results could have resulted from the use of representative samples or from resampling; (d) a laboratory calibration of the two nuclear density gages indicated very close agreement with the manufacturer's curves; (e) the nuclear instrument, when in the direct transmission position, is a much more reliable indicator of field density than when in a backscatter position and is slightly more reliable than the conventional water-balloon tests; and (f) the air-gap procedure was more reliable than the standard block only on that project believed to have a larger variation in chemical content of the soil.

•ANYONE involved in highway construction is aware of a degree of variability in the constructed product. This variability has often been the source of problems among the many professions connected with highway construction—engineers, contractors, auditors, federal agencies, administrators, and politicians. A possible solution to these problems is the implementation of construction specifications based on the variability of the constructed product. Prior to this implementation, however, the variability must be measured. It was the purpose of this research to measure the variability of percent compaction and moisture content in acceptable highway construction in North Dakota.

With the increasing rate of construction, there is a corresponding need for methods of conducting quick, accurate control tests. A method of testing that has shown promise in other areas, but prior to 1967 has not undergone thorough field tests in North Dakota, is a nuclear instrument for measuring soil density and moisture content. A second purpose of this research was to compare the accuracy of the nuclear instrument with the conventional tests that use a water-balloon for density and drying for moisture. The nuclear instrument was used in both the backscatter and direct transmission position

Paper sponsored by Committee on Soil and Rock Properties and presented at the 48th Annual Meeting.

for both standard count ratio and air-gap ratio with the intention of comparing the accuracy of each method. Before and during field use, the instruments were calibrated with soils of known densities.

DATA COLLECTION

To evaluate the true average and variability of present acceptable construction, it was necessary that (a) the sampling take place only after the compaction had been approved by the present control methods, and (b) the sample locations be selected on a purely random basis. This was accomplished by following the general sampling procedure as suggested in a research guide published in 1965 by the U. S. Bureau of Public Roads (1).

Three grading projects were selected for testing. Each was located in a major geological area of the state. Project 1 (Fingal) was located in the glaciated soil area. Project 2 (Max) was located in the end moraine area and Project 3 (Belfield) was in the nonglaciated area in the southwest corner of the state. A fourth geological area was not represented—the glacial lake area on the eastern edge of the state. With the exception of this glacial lake area, it was planned that the three projects would represent the range in variability in the state.

In each project, 50 samples were taken from approximately 100,000 cu yd of embankment. This embankment was divided into 20 units of approximately 5,000 cu yd each. Ten of these units were selected as follows: (a) using random numbers and the length of the sampling unit, the stations for the samples were selected; and (b) using random numbers along with the old and new cross sections for that station, the offset and depth coordinate locations for the samples were selected. This procedure located the sampling unit, which was a square yard of compacted roadway a lift (1 ft) in thickness.

The square yard was divided into nine 1-sq ft units, from which two were selected by random numbers for testing. At each of these 1-ft sq locations, the following tests were run: (a) in-place density using the water-balloon method; (b) in-place moisture by drying two soil samples; (c) density and moisture content readings using a nuclear moisture-density gage; and (d) sack samples taken for determination of Proctor density in the laboratory. The nuclear moisture-density gage was used in both the backscatter and direct transmission positions. In the backscatter position, the following readings were taken: flush density, flush moisture content, 2-in. air-gap density, and standard counts for moisture and density. A direct transmission reading was taken with the probe penetrating 6 in. into the soil. The results of the data collection are discussed later. As a matter of general interest, results on the first ten samples in Project 2 are reported in the Appendix.

HIGHWAY DEPARTMENT SPECIFICATIONS

The specification under which the projects were constructed are contained in this section. The standard specifications call for a compaction of not less than 90 percent of maximum dry density where the maximum dry density is based on AASHO Designation T 180. The moisture content at the time of compaction was to be not less than 75 percent of optimum nor more than that which will permit compaction to the required density. These standard specifications were in force on Projects 1 and 2.

Special provisions were enacted for Project 3 that changed the compaction and moisture requirements. All embankment 4 ft below the finished grade line of earthwork was to be constructed in accordance with the standard specifications with the exception that the embankment be compacted to not less than 85 percent of maximum dry density (T 180). The compaction of the upper 4 ft of embankment was to be not less than 95 percent of maximum dry density where the maximum dry density was determined in accordance with AASHO Designation T 99. The moisture content of the soil (in the top 4 ft of embankment) at the time of compaction was to be not less than 4 percentage points below optimum nor more than that which would permit compaction to the required density.

RESULTS OF CONVENTIONAL TEST METHODS

Research Data

The percent compaction for the three projects is shown in Figure 1. Field density was measured by the water-balloon method. Maximum dry density was obtained from a standard Proctor test on the same soil that was measured for field density. Along with the frequency distribution, the following factors are reported: $\overline{X}$, average percent compaction; σ, standard deviation of all the observations in the project; σ_a, standard deviation of adjacent observations; and n, the number of observations in the project. The small number of duplicate samples reported in Project 3 is caused in part by rainy weather that curtailed sampling, and partly by a number of samples having a different required percent compaction. The reported samples are for the top 4 ft of embankment in Project 3.

The standard deviation is a quantitative measure of the variability. It can be used to make probability statements on the variability of the data under study. For example, 68 percent of the time an observation will be within plus or minus one standard deviation of its average value, and 95 percent of the time an observation will be within plus or minus 1.96 standard deviations of its average value.

Two different standard deviations for each project result from an analysis of variance of the data. The analysis of variance separates the sampling and testing standard deviation from the total standard deviation. The sampling and testing standard deviation results from a difference in observations of the two adjacent tests and is reported here as σ_a. The standard deviation, σ, results from a difference in observations of two tests located anywhere in the project. Applying these ideas to Project 1, it can be said that 95 percent of the time the difference between a test value and the true average will be less than 6.5 percent compaction. Similarily, for two samples located anywhere in the project, 95 percent of the time the difference will be less than 8.8 percent compaction. Likewise, the percent compaction values will be less than 10.3 and 15.7 percent for Project 2 and 7.6 and 9.5 percent for Project 3.

Comparison of the required percent compaction and the average percent compaction indicated that the average is slightly less for Projects 1 and 2, and slightly more for Project 3. Another indication of the level of percent compaction is the percentage of samples with compaction less than that required. These values for Projects 1, 2, and 3 are 52, 53, and 20 percent respectively.

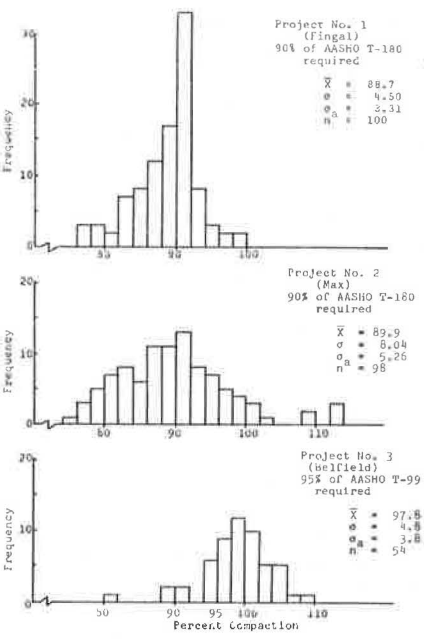

Figure 1. Percent compaction, research data.

A second variable studied is the moisture content. The moisture content alone is not of much interest, but rather is used in reference to the optimum moisture for the soil. In Figure 2 are shown the field moisture minus optimum moisture for the three projects. It can be observed that the average field moisture is near optimum for Projects 2 and 3, whereas it is about 4 percent greater than optimum for Project 1. This agrees with the field conditions during the sampling because Project 1 was temporarily stopped because of rain just prior to the research sampling. It can also be observed that the standard deviations of the differences in moisture content for the three projects are nearly equal.

HIGHWAY DEPARTMENT DATA

On completion of the testing, the State Highway Department made available their compaction control data for the three projects under study. The percent compaction data, shown in Figure 3, are for the same sections of roadway used in the research study. Recorded for each project

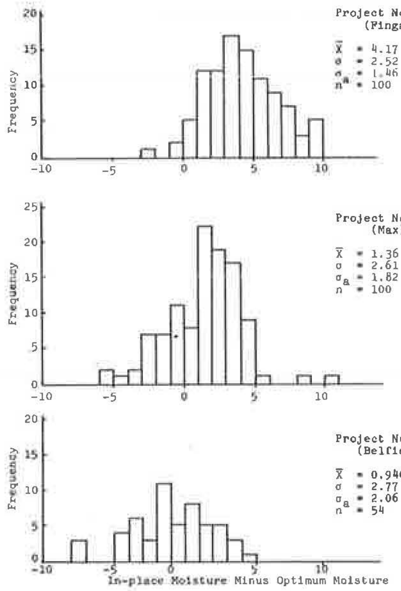

Figure 2. In-place moisture minus optimum moisture, research data.

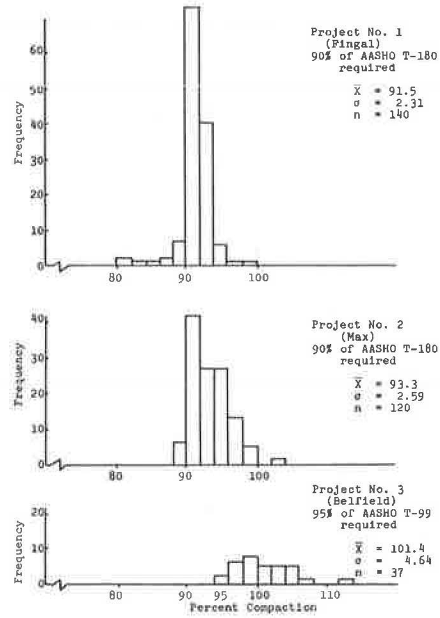

Figure 3. Percent compaction, highway department data.

are the required percent compaction, average percent compaction, standard deviation, and number of samples. Because the tests were not run in duplicate, it was not possible to identify that portion of the standard deviation due to sampling and testing of adjacent samples. The standard deviation reported represents differences in samples taken anywhere in the project. It can be noted for each of the projects that the average percent compaction is greater than the required compaction. All three projects have samples with less than the required density. However, most of these samples are within ½ percent of the required minimum. In Project 1, 6 percent of the samples were more than ½ percent below the required density.

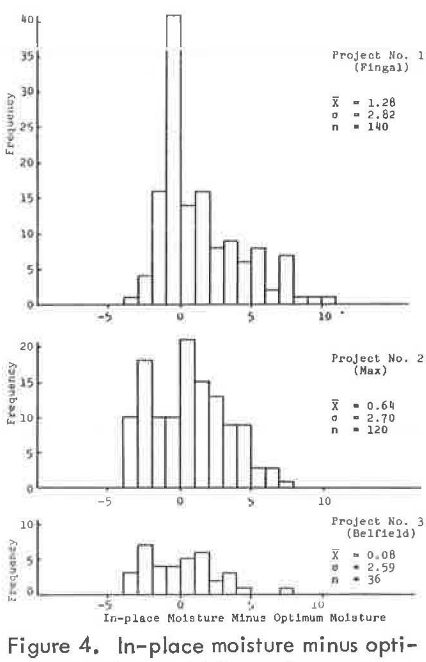

Figure 4. In-place moisture minus optimum moisture, highway department data.

Data on moisture content are shown in Figure 4. The numbers indicate the average field moisture content to be at about the optimum moisture content. Also, the standard deviation of difference in moisture content is about the same for the three projects.

Comparison of Highway Department and Research Data

It is interesting to compare the State Highway Department data with the research data from conventional tests. Figures 1 and 3 show that the average percent compaction for each project for the highway department data is about three percentage points greater than that for the research data. Also, there is a considerable difference in the shape of the frequency distribution curves.

TABLE 1

AVERAGE IN-PLACE AND MAXIMUM DRY DENSITY

Data	In-Place Dry Density			Maximum Dry Density		
	Proj. 1	Proj. 2	Proj. 3	Proj. 1	Proj. 2	Proj. 3
Highway department	106.8	112.0	105.2	117.0	120.2	103.3
Research	102.0	105.9	100.4	115.2	119.0	102.7
Highway department vs research	1.05	1.06	1.04	1.02	1.01	1.00

The research data are more symmetrical and have a larger range than the highway department data. Standard deviations for the research data are about twice that of the highway department data for Projects 1 and 2. However, the standard deviations are about equal for Project 3.

Although the percent compaction from the two sources shows considerable difference, the difference in moisture content is not readily discernible. Figures 2 and 4 indicate nearly equal standard deviations. The averages are also fairly close with the exception of Project 1, which is about 3 percent higher for the research data.

The difference in percent compaction comes from two basic sources. Because the percent compaction is a ratio of the in-place field density and the maximum dry density, the difference could result from a difference in either or both of these values. The source of this difference may be determined by comparing the average in-place density and average maximum dry density as obtained from research and highway department data for each of the projects. These values, along with a ratio of highway department values to research values, are given in Table 1. Noting that the ratios are about 1.05 for the in-place dry density and about unity for the maximum dry density, it can be concluded that the difference in average percent compaction between the highway department and research data resulted from the larger in-place dry density in the highway department data.

What are the causes for this difference in average dry density and in variability of percent compaction between the highway department and research data? Possible causes are sample location, testing technique, resampling, and the presence of a required percent compaction.

Sample locations for the research data were selected solely through the use of random numbers. That is, the bias of the technician had no effect on sample location. Sample locations for the highway department data were selected by the technician as being representative of the compaction in the area under question. This would introduce a bias if the technician were able to judge relative compaction prior to selecting a sample location. This bias could change both the mean and standard deviation of the test results. If either the denser or looser soil were sampled, this would be indicated by a lower or higher than true average dry density. If samples are being selected as representative of the average density, then the variability (standard deviation) of the test results will be lower than the true variability because samples from neither the low nor high densities will be selected.

Testing technique is another possible source of error. An attempt was made for the research personnel to use the same testing technique with the water-balloon method as was standard practice with the highway department. In particular, the base plate was staked in place and an auger was used to remove the soil. The less care and consistency exercised in performing the tests, the greater is the variability (standard deviation) of the test results. This might explain the larger standard deviation of the research test results on percent compaction; however, it is doubtful because the same degree of care would have been used on moisture content determination and in this case the standard deviation of highway department and research data are nearly identical. It should be noted that a portion of the highway department data in Projects 1 and 3 was taken with a nuclear instrument rather than the water-balloon method. This may have affected the results.

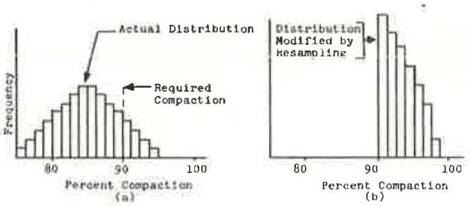

Figure 5. Effect of resampling.

Resampling can contribute to a change in both the average and standard deviation of the test results. For example, consider the case where the embankment has been compacted to an average density of 85 percent. The actual frequency distribution of the test results may be as shown in Figure 5a. If all tests indicating a compaction less than required are discarded, the resulting frequency distribution of the recorded test results will be as shown in Figure 5b. By comparing the two frequency distributions of Figure 5, it can be observed that the effect of resampling is both to increase the average percent compaction and to reduce the variability (standard deviation).

The last factor to explain this difference in test results is the effect of the presence of required percent compaction. There may be a tendency on the technician's part to record those tests with results just below passing as passing. This would especially be true if, in the opinion of the technician, the compaction was considered to be adequate and he was just looking for a test result to verify that opinion. The technician collecting the research data was not under pressure of evaluating the adequacy of the compaction, nor did he perform at that time the calculation of the densities.

Comparison With Results of Others

Before it is hastily concluded that the data presented are an indication of poor-quality construction, these results should be compared with compaction results from other areas. The prime example of expert care and control in construction (and yet not meeting the specifications) is the construction of the AASHO Road Test tract. About 9 percent of all compaction tests on embankment material were below the specified minimum requirement.

A research report from the California Division of Highways on three of their construction projects indicates that the average percent compaction was from 0.5 to 3.5 percent greater than the required compaction (2). The standard deviation ranged from 2.4 to 5.5 percent compaction, and the percentage of samples with less than the required compaction on the three projects was 9, 24, and 43 percent.

A research report from the Indiana State Highway Commission states that, for the three construction projects they tested, the average percent compaction was from 3.2 percent below to 0.6 percent above the required compaction (3). The standard deviation ranged from 4.5 to 5.7 percent compaction, whereas the percentage of samples with less than the required compaction on the three projects was 48, 68, and 72 percent.

NUCLEAR INSTRUMENT CALIBRATION

To date, nuclear instruments for measuring soil density and moisture have not been universally accepted. This is partly because of the variability in their results. It has been observed that the nuclear density values are a function of both the soil density and the chemical content. It was recently suggested that the chemical effect could be overcome by use of an air-gap procedure (4). For this reason, it was decided to calibrate the nuclear instrument for the purpose of verifying the manufacturer's curve and for establishing a curve based on the air-gap procedure.

The nuclear moisture-density instruments used were manufactured by Troxler Electronic Laboratories. Each instrument came in two parts, a gage and a scaler. The gage was Model SCM-227 surface and subsurface density and surface moisture gage. A Model 200-B transistorized portable scaler was used. As previously stated, the density gage could be used in either the direct transmission or backscatter positions.

The air-gap ratio method consists of two readings; one records the radiation count from the backscatter or direct transmission with the density gage flush with the soil, and the other records the radiation count from backscatter with the density gage supported a distance (about 2 in.) above the soil. The ratio of these two readings together with a calibration curve is used to obtain the soil densities.

Prior to actual calibration, it is necessary to establish the distance or air gap between the soil and the instrument. The procedure consists of recording and plotting the radiation count vs air gap for gaps ranging from 0 to 3 in. The air gap to use is the one with the maximum count for the most typical materials (4). By this procedure, an air gap of 2 in. was selected.

The calibration procedure was as follows:

1. Soils were selected for their range in densities from 75 to 140 pcf (pounds per cubic foot) by using sand, clay, pea gravel, and combinations thereof.
2. A rigid box open on the top and of known weight and volume was constructed (14 by 14 by 12 in.).
3. Samples were prepared in the box by placing the soil in layers to obtain uniform density. To obtain dense samples, the layers were compacted with a Proctor hammer.
4. Flush backscatter, air-gap backscatter, and direct transmission readings were taken for each instrument on the prepared soil. Readings were also taken on the standard block.
5. The box was weighed and the density computed from the known volume and weight of the box.

Using this procedure, nuclear readings were taken on as many as ten soil samples.

It was then necessary to use the data to develop equations relating the nuclear readings and soil density. An equation of the following form was developed:

$$\text{Wet density, pcf} = C_1 + C_2 \cdot f(X_1/X_2) \tag{1}$$

where C_1 and C_2 are constants determined from the regression analysis, and $f(X_1/X_2)$ is a function of the ratio of two nuclear radiation counts, X_1 and X_2.

The regression equation will give only estimates of the wet density. The accuracy of this estimate is indicated by the standard deviation of the wet density. The equation that gives the smallest standard deviation of the wet density would give the most accurate estimates. Table 2 gives the standard deviations for selected equations. The first column of Table 2 contains the functions of the two nuclear radiation counts; the terms refer to the position of the gage when the counts were made.

One can first observe that the log functions are desired over the straight ratios because they give equal or lower standard deviations. The probe (direct transmission) equations have standard deviations about half of those for flush (backscatter). The desirability of the air-gap reading in place of the standard reading is not clearly indicated because the standard deviation for the air gap is lower with instrument 205 and higher with 228. Only six samples were tested in the flush position for instrument 228; hence, the large standard deviations reported for that case are not very reliable.

To compare the manufacturer's equations with those derived above, the two lines with the data points are plotted in Figures 6 and 7. The equations are of the standard type (Eq. 1) with log (flush or probe count/standard count) inserted for $f(X_1/X_2)$. It can be noted that the laboratory curves fall almost on top of the manufacturer's curves with the exception of the flush (backscatter) equation for instrument 228.

The equations and points employing the log of the air-gap ratio are plotted in Figures 8 and 9. The instrument manufacturer did not supply curves with which to compare these equations.

After completing the field work, typical soil samples from each project were used in calibrating the two instruments. However, the resulting densities were not of

TABLE 2

STANDARD DEVIATION FOR WET DENSITY EQUATIONS

$f(X_1/X_2)$	Standard Deviations, pcf	
	Instrument 205	Instrument 228
Flush/standard	8.9	13.3
Probe/standard	6.7	4.7
Log (flush/standard)	8.9	13.0
Log (probe/standard)	5.5	3.3
Flush/air gap	8.2	14.3
Probe/air gap	5.5	5.5
Log (flush/air gap)	8.0	14.2
Log (probe/air gap)	4.5	3.9

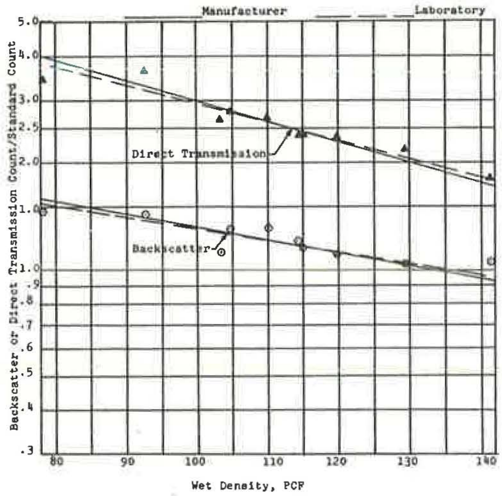

Figure 6. Calibration curves using standard count, instrument 205.

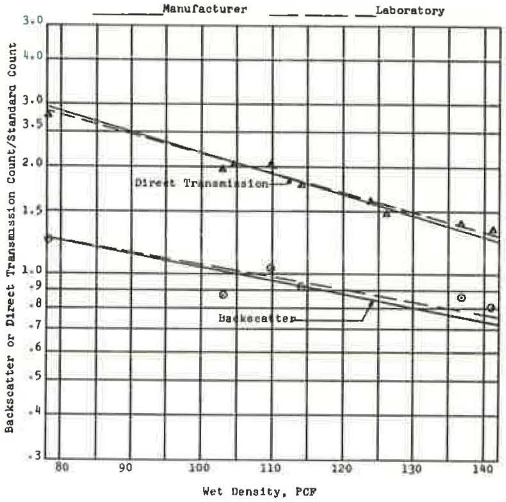

Figure 7. Calibration curves using standard count, instrument 228.

sufficient range to provide a meaningful calibration. Hence, the results are not reported.

A limited amount of calibration work was done on moisture contents. The data points, resulting regression curves, and manufacturer's curves are reported in Figures 10 and 11. For the typical range of values for instrument 205, 10 to 25 pcf, the manufacturer's and laboratory curves are nearly identical. For instrument 228, the laboratory curve gives approximately 1.5 pcf greater than the manufacturer's curves. The standard deviation from the wet density equations are 1.3 and 1.2 for instruments 205 and 228 respectively.

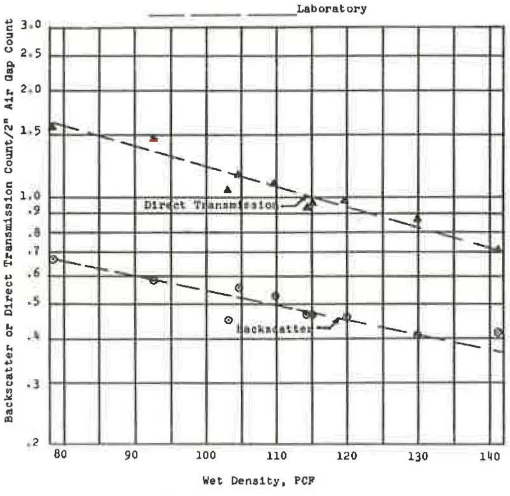

Figure 8. Calibration curves using air-gap count, instrument 205.

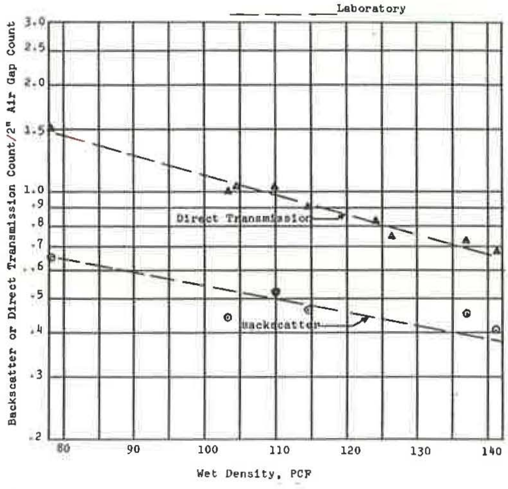

Figure 9. Calibration curves using air-gap count, instrument 228.

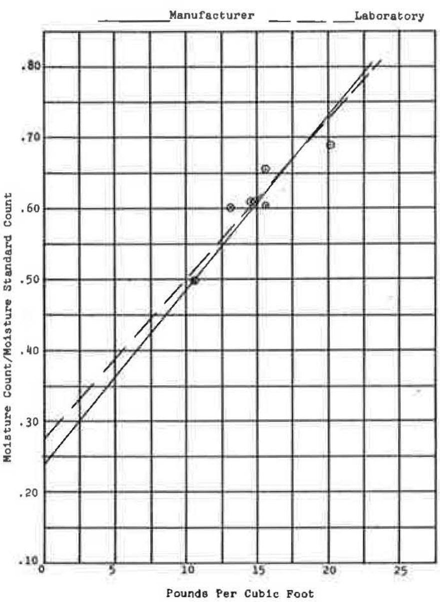

Figure 10. Moisture calibration curves, instrument 205.

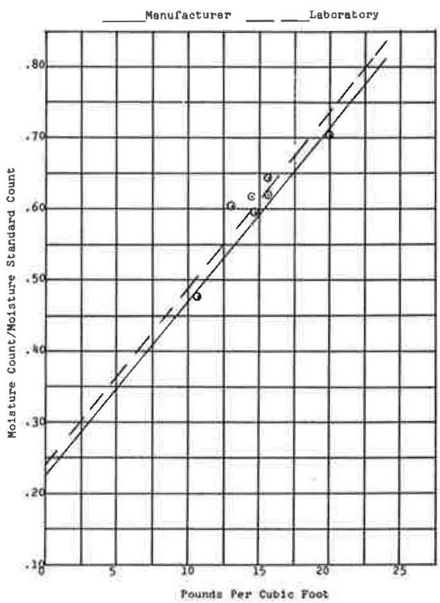

Figure 11. Moisture calibration curves, instrument 228.

RESULTS OF NUCLEAR TEST METHODS

Two sets of nuclear density data are reported for each project. One set is based on the air-gap procedure using the laboratory-prepared calibration curves. The other uses the standard block based on the manufacturer's curves. In the case of the air gap, the laboratory curves are used because the manufacturer did not supply air-gap curves. Manufacturer's curves were used with the standard block because the laboratory and

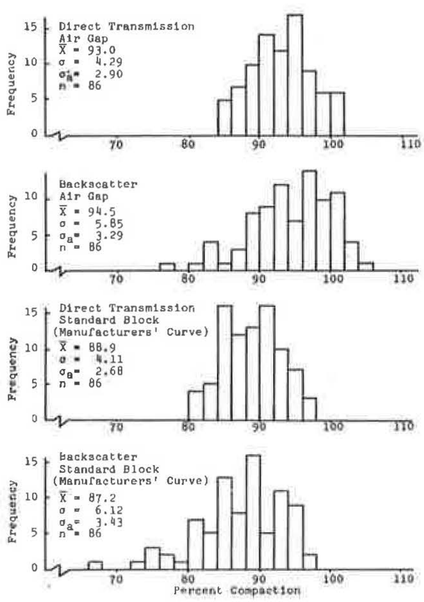

Figure 12. Percent compaction, nuclear instrument data, Project 1.

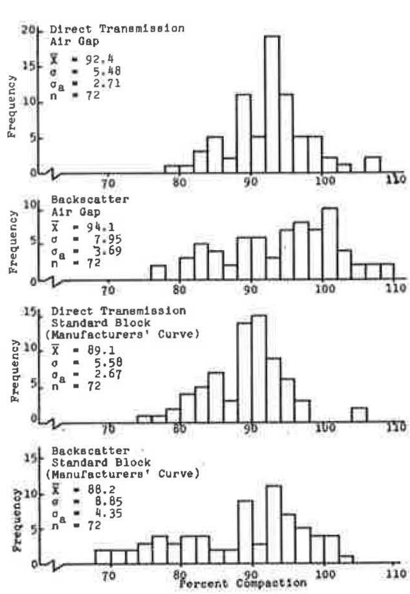

Figure 13. Percent compaction, nuclear instrument data, Project 2.

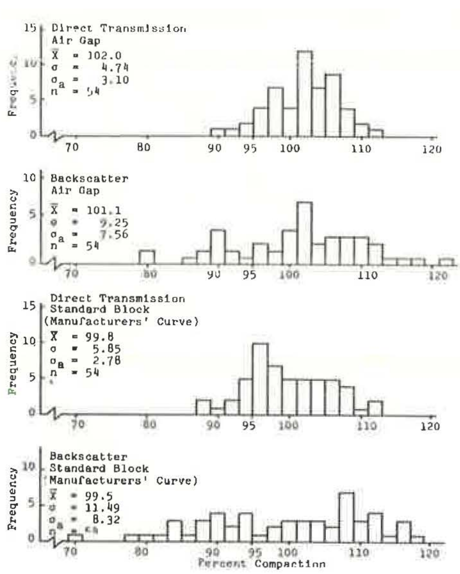

Figure 14. Percent compaction, nuclear instrument data, Project 3.

manufacturer's curves were nearly identical. All moisture contents were based on the manufacturer's curves.

The nuclear density data are reported in percent of Proctor compaction in Figures 12, 13, and 14 for Projects 1, 2, and 3 respectively. The upper two frequency distributions are for densities based on the air-gap laboratory curve whereas the lower two are for densities based on the standard block, manufacturer's curve. Each set of data contains results from direct transmission (6-in. probe) tests and from backscatter tests. Besides the frequency distribution, there is also reported the average percent compaction, $\overline{X}$; the standard deviation of all observations, σ; the standard deviation of adjacent observations, σ_a; and the number of observations, n, for each series of tests.

These data can be compared with the intent of selecting the test method that is the most reliable indicator of the percent compaction. The test method with the lowest standard deviation of test values (that is, having the least scatter) is the most reliable test method.

In comparing the backscatter with the direct transmission, it is noted that the backscatter tests all indicate larger standard deviations. In fact, they are sometimes more than twice as great as those for the corresponding direct transmission results. Hence, the direct transmission tests are more reliable.

Next, the direct transmission standard block is compared with the direct transmission air gap. For Projects 1 and 2, the standard deviations are nearly equal; however, for Project 3 the standard deviation of the air gap tests is one-fourth lower than the standard deviation for the standard block tests. Project 3 is in an area expected to contain a larger range in chemical content in the soil than in the other two projects, which explains the lower standard deviation for the air-gap tests.

A comparison of the average percent compaction for the direct transmission test methods indicates the air-gap tests to be from 2 to 4 percent higher. An instrument calibration involving more than the six to ten points used here may have resulted in a curve indicating a lower percent compaction.

Finally, a comparison is made of the nuclear densities and the water-balloon densities. The water-balloon average densities agree best with the direct transmission standard block average densities. In all three projects, lower standard deviations occur for one or both of the nuclear direct transmission tests than for the water-balloon tests. This indicates that the nuclear direct transmission tests are more reliable than the conventional water-balloon tests.

CONCLUSIONS

Through research, many heretofore unknown facts are brought to light. Probably the most significant observation from this research is the degree of variability in acceptable embankment compaction. It may be startling for someone experienced in highway construction to learn that every third test will deviate from the average percent compaction by at least 3.3 to 5.3 percent (if the maximum dry density were 100 pcf, this would be a difference in dry density of 3.3 to 5.3 pcf). Similarly, for random samples taken from anywhere in the project, the difference in percent compaction would be at least from 4.5 to 8.0 percent (the different percentages are for the different projects). Other pertinent conclusions are listed below:

1. The average percent compaction was very near the required minimum percent compaction, which led to a substantial portion of the samples with densities less than the required minimum.

2. The highway department data reported higher in-place densities and lower standard deviations on percent compaction. This could have resulted from the use of representative rather than random samples, resampling, and the presence of a required percent compaction.

3. The average maximum dry densities for the research and highway department data were nearly equal. This, no doubt, is gratifying to the highway department because the research data were obtained from a laboratory test on each sample, whereas the highway department data result from the technician matching the field soil with a description of the soil compacted in the laboratory prior to construction.

4. Results of similar research projects in other states have also shown the large variability and large percentage of samples with densities less than the required minimum.

5. A laboratory calibration of the two nuclear density gages indicated very close agreement with the manufacturer's curves. Hence, the manufacturer's curves are adequate for field use.

6. For the make of nuclear instruments used, the direct transmission position is a much more reliable indicator of field density than the backscatter position and slightly more reliable than conventional water-balloon tests.

7. The air-gap procedure was more reliable than the standard block only in that project believed to have a larger variation in chemical content of the soil.

RECOMMENDATIONS

The report brings to light a significant degree of variability in the percent compaction in present highway construction. In order for construction specifications to be effective, they must incorporate this variability in a quantitative manner. It is therefore recommended that (a) specifications containing acceptance sampling plans incorporating this variability be written, and (b) these new specifications be put in trial use along with the present specifications to familiarize the contractor and engineer's field personnel with their application.

Granted that this variability exists, one should ask, can it be reduced? If the variability is a result of causes within the control of the engineer or contractor and if the control is not too costly, then the variability should be reduced. For example, an operator hauling over the same path would add to the variability. This variability could be reduced significantly through concern by the operator. Hence, it is recommended that the contractor and engineer work together to reduce the variability where possible. Statistics-based specifications are often written to provide incentives for the contractor to reduce the variability.

Two methods of sample selection have been reported here—one using representative samples, the other using random samples. It is only through random samples that one can obtain an unbiased estimate of the actual field density. It is recommended that sample locations be determined through the use of random numbers. (This change may be best made when and if the statistics-based specifications are employed.)

On the question of quick field measurements, it is recommended that field moisture and density measurements be made with a direct transmission nuclear moisture-density instrument. It gives more accurate results than the present water-balloon method or the backscatter nuclear moisture-density instrument. The manufacturer's curve is recommended for use with the nuclear instrument. A calibration curve based on the air-gap procedure is advantageous in the nonglaciated areas of the state.

ACKNOWLEDGMENTS

This research project was sponsored by the North Dakota State Highway Department together with the U. S. Bureau of Public Roads. The opinions, findings, and conclusions in this publication are those of the author and not necessarily those of the North Dakota State Highway Department or the Bureau of Public Roads.

A sign of a mature administration is one that will permit an independent agency to examine and report on its construction control procedures. For this, the Highway Department is to be complimented.

REFERENCES

1. The Statistical Approach to Quality Control in Highway Construction. U. S. Bureau of Public Roads, April 1965.
2. Jorgenson, Frank R., and Watkins, Robert O. Compaction—Myth or Fact. California Department of Highways, Sacramento, June 1965.
3. Williamson, T. G., and Yoder, E. S. An Investigation of Compaction Variability for Selected Highway Projects in Indiana. Indiana State Highway Commission, West Lafayette, Dec. 1967.
4. Rostron, J. P., Schwartz, A. E., and Brown, F. B. Air Gap Procedure for the Measurement of Surface Density by Gamma Ray Backscatter Technique. Highway Research Circular 44, Aug. 1966.

Appendix

TYPICAL TEST RESULTS,
FIRST TEN SAMPLES ON PROJECT 2

Sample No.	Moisture, Drying-Can 1	Moisture, Drying-Can 2	Moisture, Drying-Average	Moisture, Nuclear (Density from Direct Trans)	Moisture, Nuclear (Density from Backscatter)	Optimum Moisture AASHO T-180	Dry Density Balloon	Dry Density Nuclear Direct Trans.	Dry Density Nuclear Backscatter	Dry Density Standard, AASHO T-180	Percent Compaction Balloon	Percent Compaction, Direct Transmission	Percent Compaction, Backscatter
MO1A	11.0	15.7	13.4	14.8	14.8	12.0	114.5	113.0	113.4	122.0	93.9	92.7	92.9
MO1B	11.2	10.8	11.0	13.8	14.0	12.0	117.3	116.0	114.5	122.0	96.1	95.1	93.8
MO2A	12.9	12.5	12.7	12.9	13.6	11.0	98.7	105.2	100.0	124.0	79.6	84.9	80.6
MO2B	15.4	12.7	14.1	13.1	13.5	11.0	103.7	106.1	102.9	124.0	83.2	85.6	83.0
MO3A	14.6	13.9	14.3	16.0	14.8	13.0	105.9	102.4	110.5	118.0	89.8	86.7	93.6
MO3B	13.7	12.5	13.1	16.0	16.6	13.0	101.0	102.1	98.7	118.0	85.6	86.5	83.7
MO4A	15.0	15.6	15.3	17.5	15.8	12.5	107.6	102.9	113.7	121.0	89.0	85.0	93.9
MO4B	13.8	13.8	13.6	15.0	14.5	12.5	119.3	110.1	114.1	121.0	98.6	91.0	94.3
MO5A	11.4	10.8	11.1	14.4	14.0	13.0	118.4	114.7	118.6	121.0	97.8	94.8	98.0
MO5B	10.8	12.0	11.4	13.6	14.2	13.0	120.0	117.2	112.7	121.0	99.2	96.9	93.1
MO6A	15.6	12.3	14.0	15.2	16.1	13.5	109.0	106.8	101.1	118.0	92.4	90.5	85.7
MO6B	16.3	15.3	15.8	10.9	16.6	13.5	106.3	107.2	102.4	118.0	90.1	90.8	86.8
MO7A	13.7	12.7	13.7	16.8	15.9	14.0	106.0	99.2	104.7	117.0	90.6	84.8	89.5
MO7B	13.7	12.8	13.3	15.6	14.6	14.0	97.9	104.1	111.6	117.0	83.7	89.0	95.4
MO8A	16.7	17.0	16.9	16.3	16.1	13.0	105.7	110.5	112.0	120.0	88.1	92.1	93.4
MO8B	15.1	14.8	15.0	15.9	16.2	13.0	113.4	109.6	107.7	120.0	94.5	91.3	89.7
MO9A	12.8	13.5	13.2	15.6	14.3	12.0	112.9	108.0	117.7	121.0	93.3	89.3	97.3
MO9B	13.9	14.1	14.0	10.5	15.1	12.0	114.8	102.2	111.4	121.0	94.9	84.5	92.1
M10A	9.6	10.5	10.1	12.3	13.5	12.0	109.9	108.6	99.2	120.0	91.5	90.5	82.7
M10B	11.2	11.0	11.1	12.8	15.0	12.0	108.9	107.3	91.1	120.0	90.8	89.4	75.9

Control and Acceptance of Aggregate Gradation by Statistical Methods

WILLIAM H. MILLS, William H. Mills and Associates; and
OREN S. FLETCHER, South Carolina State Highway Department

The purpose of this investigation was to develop a procedure for control and acceptance of gradations of aggregates by statistical methods. Because the South Carolina State Highway Department is most interested in gradation of aggregates when they are used, a procedure for control and acceptance was developed for "as-used" samples. This procedure was tested on 14 projects. As refined, the system provides that the size of a lot shall be the quantity required for 1 mile of 24-ft wide roadway or 100 cu yd of structural concrete. Five random samples per lot are specified. For analysis of the data, the "standards given" control chart technique is used. The central value is the desired average of the gradation specified. Upper and lower control limits are established using standard deviations previously determined. When test values exceed control limits, the delivered price for the lot will be adjusted.

In this study, analysis of variance showed variance resulting from the testing process significant in 98.8 percent of results analyzed, variance between batches in 30 percent, and variance within batches in 56 percent. A study of the testing process was recommended with a goal of reducing this variance. A limited series to explore the effect of the size of sample on gradation test results showed no definite preference for a certain size sample.

•IN HIGHWAY CONSTRUCTION, aggregates composed of rock fragments form the bulk of the materials in cement concrete, bituminous concrete, base mixtures, and subbase mixtures. The properties of these mixtures are directly influenced by the aggregates used in them. Durability may be greatly affected by the chemical and mineral composition. Workability and strength are affected by the shape, size, and gradation of the particles.

The chemical and mineral composition and shape of aggregate particles are usually fairly constant at any one source. The size of particles and gradation (proportions of different sizes) can be and are varied by crushing and screening to obtain properties of workability and strength desired for a certain mixture. Ideally, the gradation should be constant, but variations are always encountered in any mixture containing particles of different sizes. When aggregate consisting of different sizes of particles is handled or moved, segregation or separation of the sizes occurs. The tendency to segregate increases when the size of particles is large or when the range in sizes is large.

The gradation of a certain lot of aggregate is usually determined by obtaining a sample from the lot and separating the particles on appropriate sieves. The problem of obtaining this sample has always caused testing engineers much concern. Traditionally, the attempt has been to select a "representative" sample, one that will indicate the

Paper sponsored by Committee on Quality Assurance and Acceptance Procedures and presented at the 48th Annual Meeting.

average gradation of the particles in the lot. How closely this goal is attained depends on several factors, such as the experience of the one who obtains the sample, the maximum size of the particles, the different sizes of particles present, the segregation of the sizes, and the quantity of material included in the sample. Because of the many possibilities for variations in a representative sample, there is often controversy regarding the accuracy of the results.

It is the practice of the South Carolina State Highway Department to base acceptance of aggregate gradation on samples obtained from stockpiles because there is usually no other feasible sampling point. Such samples often fail to meet gradation requirements. When a sample does not conform to specification limits, the usual policy is to select two check samples and base acceptance on the results of the three samples.

The difficulties in obtaining a representative sample from a stockpile, the emphasis on no-deviation compliance with specifications, and the desire to develop a more practical and realistic procedure for control and acceptance of aggregate gradation were factors that influenced the South Carolina State Highway Department to undertake this research project. The Bureau of Public Roads agreed to cooperate in this project and to defray part of the cost with H. P. R. funds.

The State Highway Department did the sampling and testing with a crew of two men. Sampling was done at production sources and at projects. All testing was done at the central laboratory at Columbia, South Carolina.

The firm of William H. Mills and Associates was retained to do the detailed planning for this work, to maintain coordination with Highway Department personnel in performing field sampling and testing, to tabulate and analyze data, to prepare the sampling plan and the procedure for obtaining random samples, to prepare a tentative system for control and acceptance based on statistical concepts, to continue coordination with field forces during the testing of the tentative procedure, to review and refine the procedure, and to prepare model requirements for specifications to use the system.

The research project was conducted in four parts as follows:

Phase I — Determination of statistical parameters for gradations of typical aggregates used by the department;

Phase II — Preparation of tentative procedures for random sampling and acceptance of gradation of aggregates;

Phase III — Field testing and refining the tentative procedures; and

Phase IV — Preparation of models for specifications to utilize the system.

PHASE I—DETERMINING STATISTICAL PARAMETERS

Survey of Plants and Stockpiling

For background material and in order to become familiar with the methods of handling, loading, and sampling aggregates at the various sources, a survey was made of commercial plants supplying aggregate to the department. Particular attention was given to the methods of combining different sizes to obtain a specified gradation and to the methods of loading trucks and railroad cars, especially in regard to segregation of different sizes. The locations and procedures for obtaining and testing control samples were noted. Methods for stockpiling aggregate at project sites were observed.

Following this survey, two short pilot studies were conducted to investigate procedures for obtaining random samples and to develop initial information in the variations in gradation that could be expected.

Pilot Studies

The first study was conducted on crushed stone, $1\frac{1}{2}$ in. to No. 8, used in class B concrete for curb and gutter work. Samples were obtained at the source by passing a metal box through the stream of aggregate as it flowed from the loading bin to a railroad car below the loading platform. Samples were obtained during unloading from the discharge of a belt by passing a 5-gallon bucket through the stream of material as it dropped to the stockpile. The aggregate as used (actually incorporated into the concrete) was also

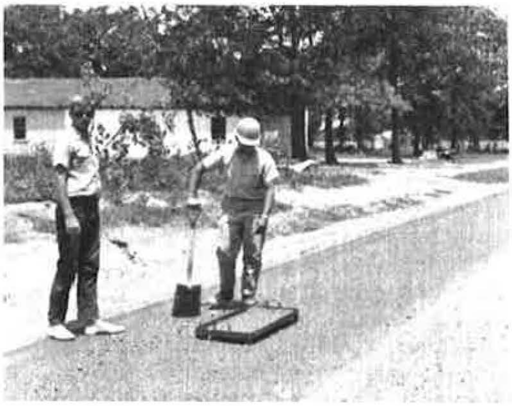

Figure 1. Tray for sampling surface treatment aggregate as used.

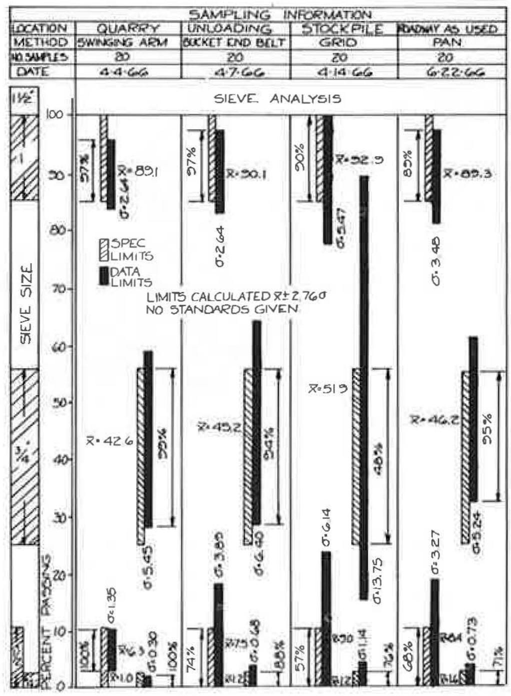

Figure 2. Aggregate No. 3, Weston and Brooker Co., Cayce, S.C., sampling information.

sampled in a similar manner as it was being delivered by chute from the mixer truck into the forms. Randomization in the selection of samples was attained by using random numbers to determine the ton or time at which the sample would be drawn. Twenty random samples were obtained at each location from this shipment of approximately 800 tons.

The second study was on 3,000 tons of crushed stone, $1\frac{1}{2}$ to $\frac{1}{2}$ in., used in surface treatment. Twenty random samples each were obtained at the source, during unloading, from the stockpile, and as the material was being placed on the roadway. Randomization was attained by using random numbers to determine the ton or time at which the sample would be drawn.

Samples at the source were obtained with a removable box mounted on a swinging arm that was passed through the stream of material as it was discharged into a railroad car. This material was unloaded by discharging it through gates onto a conveyor belt. Samples were obtained from the end of this belt by passing a bucket through the stream of aggregate.

The aggregate was deposited in the stockpile by truckloads in such a way that a succeeding load would overlap the previous load by about one half and leave a well-defined small cone from each dump. The cone to be sampled was determined by random numbers. The exact location from which the sample was drawn was then determined by random numbers.

The device developed for sampling the aggregate as used consisted of a heavy metal pan, 18 by 36 by 2 in. (0.5 sq yd in area), mounted on short supports to hold the pan above the surface of the asphalt. This pan was placed on the roadway ahead of the spreader truck at a randomly preselected location (Fig. 1). The data for these samples are shown in Figure 2.

Aggregates for Concrete Pavement

The study was continued at a concrete paving project. At this plant, aggregate in three separate sizes was fed to the weighing bin by separate belts operating from under small bins that were filled by a front end loader. The conveyor belts from these loadings bins to the weighing bin were stopped each time the hopper above the weighing bin was full. Thus, it was practical to obtain samples of the separate sizes of aggregate from stopped loaded belts (Fig. 3).

Figure 3. Sampling from a stopped belt.

Randomization for the five samples per lot (quantity used in a day) was attained by using random numbers to select the exact time to draw the sample.

Sampling of the concrete from the roadway was done by placing a container at a randomly preselected location on the subbase and having this container filled as the spreader passed over the area.

Samples were obtained from the outer layer of the large stockpiles by dividing the outer surface of the stockpile into areas approximately 10 by 10 ft and subdividing these 1-ft squares. The exact squares for sampling were determined by random numbers.

Aggregate at another concrete paving project was sampled later in the year. At this plant, sizes of aggregates were weighed in individual hoppers and fed by belts to a common bin from which they were moved to the mixers. By switching the plant to manual operation, it was possible to obtain samples from the individual loaded belts after the aggregate had been weighed.

Aggregates for Concrete for Structures

A series of tests was made on aggregate, 1½ in. to No. 8, used on a bridge project. The size of the lot to be sampled was chosen as the quantity used in a normal day's placement of concrete on the bridge deck (about 100 cu yd). Samples of aggregate were obtained at the source from a stopped loaded belt, and as-used samples from a stopped loaded belt at the batch plant. Samples of concrete were obtained from the chute as concrete was delivered to the forms. Randomization for the five samples was based on the time required for the day's production of concrete. The data are shown in Figure 4.

Aggregates for Surface Treatment

The original sampling plan for surface treatment aggregates, 1½ to ½ in. and ⅜ in. to No. 16, was to obtain five random samples from a shipment, considered as one lot, at the source, at the stockpile, and as the material was being used. This program could not be followed, however, because most of the aggregates for the current construction season had already been stockpiled at project sites. The procedure followed in most instances was to sample the aggregate in the stockpile considered as a lot, and then sample that material as it was being placed on the roadway. Later, samples were obtained at the source by the method used by the producer. Such samples give typical results for the source, but they are not comparable directly to the material sampled in the stockpile or as used.

The procedure described in the pilot study was followed to obtain the stockpile and the as-used random samples.

Replicate Sampling

To obtain data for the analysis of variance, replicate samples were obtained from two lots of each aggregate size at each sampling point. Replication requires taking two samples at essentially the same time. Where samples were taken from a stockpile or from a stopped conveyor belt, the replicate samples were taken adjacent to each other. Where the material was being discharged from a bin or from a conveyor belt, the two samples were obtained within the shortest time practicable. For roadway samples, two devices were placed side by side.

Testing

All tests (sieve analysis) were performed at the central laboratory in Columbia, S.C. A Gilson shaker and one set of sieves were used throughout. A Gilson aggregate splitter was used to reduce samples to test portions when the sample size was greater than the normal capacity of the shaker, as was usually the case.

The actual performance of the sieve test was routine. The duration of the shaking varied somewhat with the quantity of fine material in the sample. Shaking was continued until there was no visible evidence of material passing the two smallest screens. Smaller size aggregates, such as $^3/_8$ in. to No. 16, were reduced to test size in a sand splitter and sieving was performed in a Rotap.

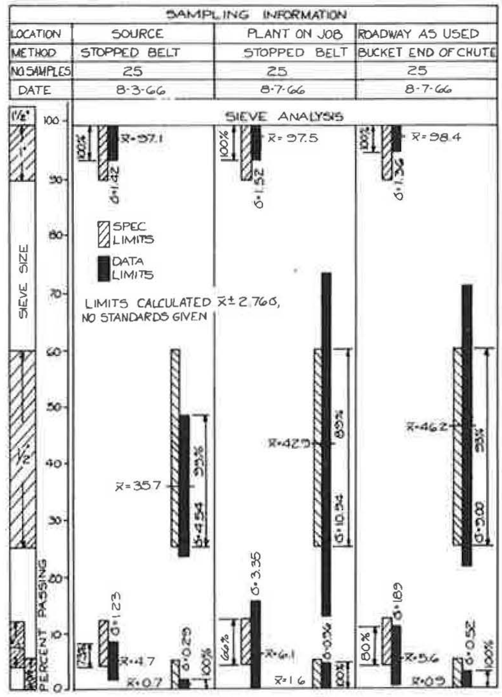

Figure 4. Aggregate No. 4, Becker County Sand and Gravel Co., Marlboro, S.C., sampling information.

Analysis of Data

Data in Phase I included gradation test results of 175 samples obtained at the source, 120 samples obtained during unloading on the project, 320 samples from stockpiles, and 400 samples taken as the material was being used. Coarse, intermediate, and fine gradation of aggregates were included, as well as crushed aggregate produced at five sources and gravel produced at two sources.

The data for the samples from each location were analyzed to obtain mean values ($\overline{X}$) and standard deviations (σ) for each sieve in the gradation series. These data were used in calculating upper and lower control limits on the basis of no standards given ($\overline{X} \pm 2.76\, \sigma$) for each sieve. Typical results are summarized and compared graphically with specification limits in Figures 2 and 4. In each case the percentage conforming with specifications is shown. The statistical values for samples in Phase I are summarized in Tables 1 and 2.

Conclusions

Based on experience and analysis of the data in Phase I, the following conclusions were derived:

1. In most cases, samples obtained at the source conform to specifications.
2. Samples obtained from stockpiles on the project show wide variations and frequently do not meet specifications.
3. Samples obtained as the aggregate is used show less variation than the stockpile samples but somewhat larger variation than the source samples.
4. Source samples show less variation than others because they are obtained from the stream of aggregate after the several sizes have been measured in proportions to give the gradation desired and before discharge or handling that would cause segregation. At some producers, source samples are obtained from a stopped loaded belt; at others, by cutting through the stream of material as it is discharged from the loading belt.
5. Sampling from a stockpile always involves difficulties because of the segregation of different-sized particles. Also, in most cases, only the material near the surface of the pile can be included in the sample.

TABLE 1

SUMMARY OF STATISTICAL VALUES OF AGGREGATES FOR CONCRETE—PHASE I

Sample	σ	$\bar{X}$	%	σ	$\bar{X}$	%	σ	$\bar{X}$	%	σ	$\bar{X}$	%
Aggregate No. 4, Crushed Stone, Columbia												
Sieve size		1 in.			½ in.			No. 4			No. 8	
Source	1.11	97.8	100	2.95	30.3	96	1.44	7.8	100	0.46	1.6	100
Unloading	1.80	97.6	100	6.95	36.0	94	3.78	8.4	71	0.95	2.1	100
Concrete	4.72	95.6	88	11.80	32.9	74	2.67	4.7	61	0.35	0.2	100
Aggregate No. 4, Gravel, Marboro												
Sieve size		1 in.			½ in.			No. 4			No. 8	
Source	1.42	97.1	100	4.54	35.7	99	1.23	4.7	73	0.29	0.7	100
Belt (as used)	1.52	97.5	100	10.94	42.9	89	3.35	6.1	66	0.96	1.6	100
Concrete	1.36	98.4	100	9.00	46.2	93	1.89	5.6	80	0.52	0.9	100
Aggregate 1A, Crushed Stone, Columbia and Cayce												
Sieve size		2 in.			1½ in.			1 in.			½ in.	
Stockpile	7.80	91.9	56	15.77	43.9	66	5.57	4.8	97	1.84	0.6	99
Belt (as used)	3.31	94.6	92	6.03	58.3	97	4.75	13.7	61	2.29	3.9	69
Aggregate 4M, Crushed Stone, Columbia and Cayce												
Sieve size		1 in.			½ in.			No. 4			No. 8	
Stockpile	3.60	94.4	89	14.27	39.3	78	5.49	8.5	54	2.02	2.6	100
Belt (as used)	1.56	95.5	100	7.06	46.7	97	3.15	13.1	95	1.35	5.0	100
Aggregate No. 20, Sand, Two Sources Mixed												
Sieve size		No. 4			No. 16			No. 50			No. 100	
Belt (as used)	0.03	99.9	100	6.46	76.4	100	2.32	16.7	100	0.68	3.8	100
Aggregate No. 20, Sand, One Source												
Sieve size		No. 4			No. 16			No. 50			No. 100	
Belt (as used)	0.00	100	100	0.86	94.2	100	1.44	9.1	100	0.68	2.4	100
Mixed Aggregates from Concrete												
Sieve size		2 in.			1 in.			½ in.			No. 4	
Concrete	1.91	98.4	100	3.55	54.5	100	4.23	23.6	100	2.23	6.7	97
Belt (calculated)			96			84			98			88
Aggregate No. 1B, Crushed Stone, Augusta												
Sieve size		1½ in.			1 in.			¾ in.			⅜ in.	
Belt (as used)	3.24	90.5	54	5.02	33.8	100	2.17	9.7	98	1.07	1.1	99
Routine	1.71	92.3	91	1.55	37.5	100	1.29	10.8	100	0.28	1.6	100
Aggregate No. 6M, Crushed Stone, Augusta												
Sieve size		¾ in.			⅜ in.			No. 4			No. 8	
Belt (as used)	1.76	91.6	82	2.54	24.4	96	0.84	7.0	100	0.40	2.3	100
Routine	0.92	91.5	95	3.81	29.9	100	0.70	5.7	100	0.24	1.2	100
Aggregate No. 20, Sand												
Sieve size		No. 4			No. 16			No. 50			No. 100	
Belt (as used)	0.0	100	100	3.18	78.4	100	1.42	14.0	100	0.41	2.5	100

Note: $\bar{X}$ = mean of data percentage passing; σ = standard deviation; % = percentage of results within specification limits.

6. For sampling at concrete plants, it is often practical to obtain as-used samples from a stopped loaded belt en route to the weighing bin or mixer. In cases where belts are not used for transporting the aggregate, as-used samples can be obtained by dropping the material for the sample from the storage or weighing bin into a suitable container.

7. A suitable device for sampling surface treatment aggregate as used consists of a tray 0.5 sq yd in area. This tray is placed on the roadway at preselected random locations.

8. It is interesting to note that for all sizes and gradations of aggregates included in this study, the largest variations (sigmas) occur at the third sieve smaller than the maximum size. In the summaries, data for the maximum size (100 percent passing) are not shown; therefore, the second size shown is actually the third size in the gradation series.

Analysis of Variance

To obtain data for the analysis of variance, the following plan for replicate sampling was adopted:

1. The location for sampling was selected, i.e., stockpile or as used.
2. For a certain lot of aggregate, five units or batches were selected by random numbers for sampling. Thus, between-batch variance, σ_b^2, could be evaluated.

TABLE 2

SUMMARY OF STATISTICAL VALUES OF AGGREGATES FOR SURFACE TREATMENT—PHASE I

Sample	σ	$\overline{X}$	%	σ	$\overline{X}$	%	σ	$\overline{X}$	%	σ	$\overline{X}$	%
Aggregate No. 3, Crushed Stone, Cayce												
Sieve size	1 in.			3/4 in.			1/2 in.			No. 4		
Source	2.15	89.1	97	5.45	42.6	99	1.35	6.3	100	0.30	1.0	100
Unloading	2.64	90.1	97	6.40	45.2	94	3.89	7.5	74	0.68	1.2	88
Stockpile	5.47	92.9	90	13.75	51.9	48	6.14	9.0	57	1.14	1.2	76
As used	3.48	89.3	89	5.24	46.2	95	3.27	8.4	68	0.73	1.6	71
Aggregate No. 3, Crushed Stone, Rion												
Sieve size	1 in.			3/4 in.			1/2 in.			No. 4		
Source	2.60	91.6	99	8.50	42.7	91	2.18	5.4	98	0.36	1.36	96
Unloading	1.62	93.1	100	5.81	40.5	99	2.18	6.2	96	0.65	1.6	75
Stockpile	3.81	92.2	88	13.58	51.6	45	9.75	13.9	34	2.06	2.4	28
As used	3.98	92.0	96	7.99	41.7	93	2.88	6.2	91	0.63	1.5	81
Aggregate No. 3, Crushed Stone, Stoney Point												
Sieve size	1 in.			3/4 in.			1/2 in.			No. 4		
Source	2.25	86.5	75	6.34	31.0	83	2.26	3.9	99	0.51	1.5	84
Stockpile	2.92	90.8	98	8.14	39.0	93	5.12	8.7	60	1.98	3.4	34
As used	3.05	88.0	83	5.64	30.9	85	3.53	6.8	82	1.89	3.1	36
Aggregate No. 3, Crushed Stone, Augusta												
Sieve size	1 in.			3/4 in.			1/2 in.			No. 4		
Source	1.61	92.3	100	5.93	29.2	76	1.95	4.5	100	0.67	1.0	94
Unloading	3.60	92.8	92	8.70	39.4	91	4.01	8.3	67	0.38	1.2	98
Stockpile	4.58	89.0	81	7.25	32.6	85	3.26	5.9	89	0.84	1.2	82
As used	4.23	91.3	93	12.78	41.8	76	6.11	10.0	50	0.76	1.6	71
Aggregate No. 3, Gravel, Hagood												
Sieve size	1 in.			3/4 in.			1/2 in.			No. 4		
Source	2.34	93.2	100	5.14	40.5	100	2.29	4.7	99	0.45	0.6	100
Stockpile	2.23	95.4	100	9.78	46.6	79	5.36	6.3	75	1.04	1.2	77
As used	3.53	92.6	98	6.63	39.6	98	3.92	6.7	80	1.95	1.8	52
Aggregate No. 9, Crushed Stone, Rion												
Sieve size	3/8 in.			No. 4			No. 16			No. 100		
Source	4.79	85.0	85	7.16	39.6	92	0.94	3.2	100	0.13	0.5	100
Unloading	4.20	90.2	99	7.92	40.2	89	2.54	3.9	79	0.39	0.8	100
Stockpile	5.57	89.6	96	9.58	40.9	81	1.80	3.7	90	0.34	0.7	100
As used	4.42	89.3	98	5.99	40.1	95	1.78	3.9	88	0.37	0.7	100
Aggregate No. 9, Crushed Stone, Stoney Point												
Sieve size	3/8 in.			No. 4			No. 16			No. 100		
Source	7.84	78.8	44	7.54	29.0	88	1.62	4.3	86	0.90	1.3	79
Stockpile	5.18	85.4	85	7.45	34.1	97	2.78	6.6	41	0.75	2.1	46
As used	3.68	82.8	83	7.22	31.4	94	1.89	5.0	70	0.52	1.2	93

σ = standard deviation. $\overline{X}$ = mean of data, percentage passing. % = percentage of results within specification limits.

3. At each of these batches or units, two separate samples were drawn: adjacent to each other if sampling a stopped belt or stockpile; within the shortest time practical if sampling a discharge stream; or with trays or containers side-by-side if sampling the roadway. This sampling procedure provided data for the determination of within-batch variance, σ_W^2.

4. Each of these samples was reduced by splitting into two test portions. This procedure provided data for the determination of variance due to testing, σ_t^2.

This study included 84 individual results. The variance of testing was significant in all except one instance or in 98.8 percent of the results. The variance between batches was significant in 30 percent of the results, and the variance within batches was significant in 56 percent of the results.

Investigation of Sample Size

Investigation of the effect of the size or quantity of material in the sample on the result of the gradation test was not contemplated in the outline for this research project. It was undertaken because at one plant standard deviations in the results from a few samples consisting of approximately 20 lb, compared with those of samples of approximately 40 lb, showed much smaller variations in the 40-lb samples.

The quantity of aggregate to be included in a sample for a gradation test has been a concern of testing engineers for many years. Standard test procedures of the American Association of State Highway Officials and the American Society for Testing and

Materials specify the size or quantity for a sample varying with the maximum-size particles in the mixture of aggregate. These requirements have not been changed in many years. It is logical that the quantity for a sample should vary with the size of particles because the inclusion or loss of one large particle in a small sample could change the percentage values. However, there is a natural and practical tendency for a sampler to take as small a quantity of aggregate as he can justify because of the work involved in the sampling and handling processes.

Arrangements were made with a gravel producer to stop the belt during the loading of a railroad car and to allow time for taking the samples desired. Five replicate series of samples were obtained. Each series had samples of approximately 200, 100, 50, 25, and 12.5 lb. The loading belt at this plant was 125 ft in length. Each series of samples required material from about 16 ft. The five sections to be sampled were determined by random numbers.

Several sets of tests were performed on these samples as follows:

1. Routine Test, Mechanical Splitter—The larger samples were reduced to test size in a Gilson mechanical splitter in the standard manner. The first test thus obtained is the one normally used in routine testing. However, each large sample was reduced to separate test portions. Thus, eight results were obtained for each 200-lb sample, four for each 100-lb sample, two for each 50-lb sample, one for each 25-lb sample, and one for each 12.5-lb sample.
2. Quartering—To compare results obtained by splitting samples in the mechanical device with those obtained by reducing samples to test size by the quartering method, the larger samples (200 lb, 100 lb) were re-mixed and then quartered by the standard method to test size. Thus, eight results were obtained for the 200-lb samples and four results for the 100-lb samples.
3. Layer Samples—To compare size of samples from aggregate spread in a relatively thin layer, such as in an ideally formed stockpile, the material in all these samples was spread into a layer approximately 6 by 6 ft by 6 in. thick. The surface of this layer was divided into 36 one-ft squares. Five 12.5-lb samples were obtained from the central portion of 5 randomly selected squares. A cylinder or ring having a diameter large enough to yield the quantity desired was pushed into the layer and the material inside was withdrawn as the sample. After testing, the material was returned to the square from which it was taken. Next, five 25-lb samples were obtained from 5 randomly selected squares using a similar procedure. Similarly, five 50-lb, five 100-lb, and five 200-lb samples were obtained and tested. For the large samples, the randomly selected square was used as the central portion for the sample with the remainder coming from adjacent squares.

The test data were transferred to punch cards that were put through a computer programmed according to statistical methods to determine mean values and components of variance, with the following results:

1. Mean testing—There is no significant difference in means (5 percent level) for the different size samples from 12.5 to 200 lb. There is more variation within large samples than between samples of different sizes. This conclusion is true no matter if the sample is split using quartering or the splitting method. This conclusion is subject to some controversy, however, because one of the basic assumptions of analysis of variance may have been violated, i.e., the assumption that the variances are equal.
2. Variance testing—Quartering produces significantly greater variability than splitting on the 100- and 200-lb samples. Based on Cochran's test and the F ratio test, the proper choice of sample size is not clear; i.e., it is not clear if one obtains less variability by testing 200 lb or 12.5 lb of material. Based on the results of the computer output of variances and Cochran's test and the F ratio test at the 5 percent level, and the practicality of testing, the 25-lb sample seems to be appropriate. There is some indication, however, that the 50-lb sample may produce a more acceptable variability when one considers sample size. In no case, however, should a sample be reduced by using the quartering method.

3. Ranges in test results of the following magnitude were found in the results for the five sizes of samples.

Sieve Size	Percent
1 in.	1.1 to 3.3
½ in.	1.8 to 7.4
No. 4	1.9 to 3.3
No. 8	0.6 to 1.4

4. Samples from the stopped loaded belt show less variation than do samples from the layer.

PHASE II—TENTATIVE PROCEDURES FOR GRADATION ACCEPTANCE

The field experience and data in Phase I of this research project led to the following statements as guidelines for a system of control and acceptance of aggregate gradations:

1. The gradation of any given lot of aggregate must be controlled initially at the source. The sampling and testing can be done by the producer or by the State Highway Department. However, acceptance at the source is impractical because of the possibilities of segregation during subsequent handling before the aggregate is used. The producer should test each lot or shipment to determine conformity with specifications.
2. The chances for variations in stockpile samples are so great that such samples are entirely undependable for control or acceptance purposes.
3. Even though a lot of aggregate may be graded within specifications as it is loaded, segregation may occur during subsequent handling, and samples obtained from the lot as the material is used will fail to meet specifications.
4. The State Highway Department wants and expects aggregate to meet gradation requirements when it is used. Therefore, as-used samples of the aggregate are most pertinent to a realistic acceptance plan.

From this background, tentative procedures for random sampling and tentative procedures for acceptance of gradations of aggregates were developed. The details are not given here because of revisions later.

PHASE III—FIELD TESTING OF TENTATIVE PROCEDURES

The tentative procedures for obtaining random as-used samples and the tentative method for determining acceptance of gradation test results were tested under regular field conditions to determine the applicability of these procedures to routine operations.

The field work was planned to include various sizes and types of aggregate most commonly used. All samples were obtained by the same team of two men in order to remove the variable resulting from different operators. Samples were obtained as the aggregate was being used from loaded belts, where possible, or from the roadway. Sixty lots of aggregate were sampled on projects. Five random as-used samples were obtained from each lot. Twenty-three lots were replicated to develop data on the sources of variations in the test results. The size of the lot was considered as the quantity of aggregate used during the day on which the samples were obtained. The time or batch or station at which the samples were obtained was determined by random numbers.

Analysis of Data

Many of the results did not conform to the requirements of present specifications. The variations from specification limits were very large in some instances, and they indicate that some lots of material did not conform to specification requirements.

Study of the test data compared with the tentative procedure for acceptance disclosed that the procedure was workable but somewhat complicated. The incidence of failure to conform with control limits was somewhat higher than the failures to conform to existing specifications. The tentative tables of percentages for payment gave drastically reduced prices for payment in some instances. Thus, it became evident that revisions in the procedure were needed.

The analysis of variances showed definite indications that variance resulting from testing is the major source of variation. The variance between batches and within batches is very large in a few instances, but the data indicate that these variations are random in nature and can be expected on that basis. The constant appearance of high values for testing variance indicates that a study to try to reduce testing variance is needed.

Revised System

Based on the foregoing analysis, sone details of the system were revised. The system as revised specifies that a lot of aggregate will be considered for acceptance on the basis of the results from five random samples obtained from the lot as the aggregate is being used. The quantity of aggregate that will constitute a lot will be the quantity of aggregate used in 1 mile of 24-ft wide concrete pavement, surface treatment, or base course. For structures, the quantity of aggregate used in 100 cu yd of concrete or equivalent volume will constitute a lot. These assumptions would give a lot size of approximately 2,000 tons of each coarse aggregate for concrete pavement, 350 tons of coarse aggregate and 150 tons of fine aggregate for surface treatment, and 4,500 tons for macadam or stabilizer aggregate.

The gradation results for the five random samples from the lot will be compared with standards-given control charts for which the engineer has previously established central values and standard deviations. Upper and lower control limits for the individual samples will be determined by the formula $\overline{X}' \pm 2.33\sigma$, and for grouped data, average of five results, by the formula $(N = 5) = \overline{X}'' \pm 1.04\sigma$.

The lot of aggregate will be considered for acceptance according to the following criteria:

Case I—All results are within control limits. When, for all sieves, individual results and the average of the five results are within the respective tolerances of the upper and lower control limits, the lot will be accepted.

Case II—Individual results are out of control. When, for a certain sieve, one or more individual results exceed the tolerances of the upper or lower control limit and the average of the five results is within the respective tolerances, the payment for the lot will be adjusted according to the following procedure:

1. The percentage of excess for each individual deviation will be determined by the following formula:

$$\text{Percentage of Excess} = \frac{X - (\overline{X}' \pm 2.33\sigma)}{2.33\sigma} \cdot 100$$

Where

X = individual test result,
$\overline{X}'$ = desired average, and
σ = standard deviation.

2. The percentage for payment for each deviation will be determined according to the following schedule:

Percentage of Excess	Percentage for Payment	Percentage of Excess	Percentage for Payment
0	100	30 to 60	98
0 to 15	99.5	60 to 100	95
15 to 30	99	100 and over	90*

*The engineer will direct whether to adjust at this figure and leave in place or remove and replace.

3. The payment for the lot will be the delivered price of the aggregate multiplied in series by the percentage for payment for each deviation.

Case III—Individual results are within tolerances but the average of the five results is out of control. When, for a certain sieve, the individual results are within the tolerances of the upper and lower control limits but the average of the five results exceeds the tolerances of the upper and lower control limits for grouped data, the price for payment for the lot will be adjusted according to the following procedure:

1. The percentage of excess will be determined by the formula:

$$\text{Percentage of Excess} = \frac{\overline{X}\ (N = 5) - (\overline{X}' \pm 1.04\sigma)}{1.04\sigma} \cdot 100$$

where

$\overline{X}$ = average of five results,
$\overline{X}'$ = desired average, and
σ = standard deviation.

2. The percentage for paymant will be determined according to the following schedule:

Percentage of Excess	Percentage for Payment
0	100
0 to 15	99
15 to 30	98
30 to 60	95
60 to 100	90
100 ±	80*

*The engineer will direct whether to adjust at this figure and leave in place or remove and replace.

3. The payment price for the lot will be the delivered price of the aggregate multiplied by the percentage for payment.

Case IV—Individual results and average of five results are out of control on one sieve. When, for a certain sieve, one or more individual results exceed the tolerances of the upper and lower control limits for individual results and the average of the five results also exceeds the tolerances for grouped data, the payment for the lot will be adjusted according to the following procedures:

1. The percentage for payment for the deviations of individual results will be determined by the method given for Case II.
2. The percentage for payment for the average of the five results will be determined by the method given for Case III.
3. The percentage for payment for the lot will be the smaller of the percentages for payment as determined in 1 and 2 immediately preceding. The payment price for the lot will be the delivered price of the aggregate multiplied by the percentage for payment.

Case V—Individual results and/or average of five results exceed tolerances on two or more sieves. When, for two or more sieves of a gradation, the individual results or the average of the five results exceed the respective tolerances of the upper or lower control limits, the payment for the lot will be adjusted according to the following procedure:

1. The payment price for the deviations on each sieve will be determined by the methods prescribed for Case II, III, or IV.
2. The payment price for the lot will be the delivered price of the aggregate multiplied in series by the payment price for each sieve on which there are deviations.

Conclusions

The test results for the samples obtained in Phase III were compared with present specifications and with the system as revised. Conclusions are as follows:

1. Random as-used samples—The procedures for obtaining random samples of aggregate as the material is being used are practical for routine operations. Usually, a technician and a helper are needed for this sampling. Because plant layouts and procedures are not standardized, the exact sampling procedure at a plant must be established by an engineer who is familiar with the theoretical background of random sampling.

The lot sizes for different aggregates as recommended herein are a compromise between the desire for accuracy, the cost of the material, and the cost of inspection. For concrete pavement and structures, the frequency of sampling is about the same as now practiced. For stabilizer and macadam aggregate, the frequency is much less than now practiced because gradation of this aggregate is not critical. For surface treatment, the frequency recommended is more than now practiced, but may be justified because gradation of this aggregate is critical to the quality of the construction.

2. Present specifications—Conformity with present specifications varied from complete conformity in a few lots to a complete failure in others. The incidence of failure to conform is considered very high—20 percent of the individual results exceed limits and 33 percent of the samples fail to meet these requirements. However, there is no background of test data on as-used random samples and more variation would be expected in random samples than in representative samples. Considering that all material had been accepted, the data show that present control procedures do not ensure that the aggregates as used will conform to present gradation requirements.

3. Acceptance criteria—The revised system for control and acceptance is developed around the assumption that the aggregates sampled in Phase I and Phase III did produce acceptable results in pavement and structure. The standard deviations shown in Table 3 (Phase IV) were established from the test data on as-used samples.

Even with standard deviations established from the test data, there were many deviations outside of the control limits. These deviations are indications of large variations (nonuniformity) in the product, and show a need for effort to improve the uniformity.

4. Price adjustment—Sixty lots were sampled in Phase III; eight of these could not be analyzed because of large variations from specifications. According to criteria, adjustment in price would be due on 40 lots.

The number of these lots on which adjustments would be made according to the revised criteria is much larger than would be the case in actual practice. On a regular construction project when results of samples from a lot exceeded control limits, the contractor would be expected to make corrections so that succeeding lots would be acceptable. However, in this research work, the samples for each lot were obtained from the aggregate currently being used without regard to the results for a lot sampled previously. Therefore, no corrections were made. Typical of this difficulty were results on one project where 22 of the 24 samples contained excess fine material and on another where 19 of the 25 samples were deficient in fine material.

Also, in specifications to utilize this system, requirements can be established so that fewer instances of adjustments in price will be needed. In order to conform to present practice, the desired averages for the several gradations were set initially at the mean of present specifications. By modifying these values, slightly in most cases, the incidence of failure to conform to control limits can be greatly reduced. For example, for Aggregate No. 4 a change in the desired average for passing a 1-in. sieve from 95 to 97 percent would eliminate deviation on this sieve in 7 of the 10 lots tested. Such data were considered in preparing the drafts of models for specifications (Phase IV).

5. System for control and acceptance—The system for random sampling and control and acceptance of aggregate gradation as presented herein is practical. It will give much more accurate data on the gradation of the aggregate being used than the present method, which is based on samples from stockpiles. This system requires random samples obtained at a location as near as practical to where the aggregate is being used and where the gradation of the aggregate is the same as it is when the aggregate is incorporated into the work. This system provides a definite procedure for dealing

with lots of aggregate when test results do not conform to specification limits. The method of determining the adjustment in price is simple and easy to operate. The details of the system can be modified readily to fit new conditions and new materials without altering the basic procedures. Lot size can be changed and standard deviations can be altered with additional experience. The schedules for percentages for payment should be modified if experience indicates that the present figures do not coincide with judgment as to the effect of the deviation on the severiceability of the finished product

PHASE IV —SUGGESTED MODELS FOR SPECIFICATIONS

Models for specifications to utilize the results of this research project in regular construction are included in the unabridged final report. These specifications are based on control and acceptance of gradation determined from test results on random samples obtained as the aggregate is used or placed in the work. Generally, the gradation requirements contain a desired average value for each sieve size with tolerances or control limits for individual test results and for the average of the five results from a lot. The desired average is based on the mean of the present specifications. In a few instances, modifications have been made in order to define more exactly the gradation desired for a definite purpose or to eliminate a technical limit that would not improve the usefulness of the product. The tolerances or control limits are established by standard statistical formulas using values for variation (standard deviations) found in this research work. The desired average ($\overline{X}'$) and the standard deviations (σ) used in preparing the suggested models of specifications for each aggregate are given in Table 3.

The suggested specifications include a schedule for making an adjustment in the delivered price for a lot of aggregate when test results do not conform to the tolerances of the control limits. This schedule provides for an adjustment in the delivered price

TABLE 3

VALUES USED IN PREPARING MODELS OF SPECIFICATIONS

Sieve Size	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ
Aggregate Number	1		1A		1B		2		2A		4		4M	
$2\frac{1}{2}$ in.	100	0.50	100	0.50			100	0.50						
2 in.	97	3.00	97	3.00	100	0.50	97	3.00						
$1\frac{1}{2}$ in.			52	8.00	97	3.50			100	0.50	100	0.50	100	0.50
1 in.	52	8.00	7	5.00	37	6.50	67	8.00	85	8.00	97	2.00	95	2.50
$\frac{3}{4}$ in.					7	4.00								
$\frac{1}{2}$ in.	20	6.00	2	2.00			35	11.00	47	11.00	42	8.00	42	8.00
$\frac{3}{8}$ in.					2	2.00								
No. 4							22	9.00	22	9.00				
No. 8	1	2.00									1	2.00	5	2.00

Sieve Size	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ
Aggregate Number	3		4X		5		6		6M		9		9M	
$1\frac{1}{2}$ in.	100	0.50												
1 in.	95	4.00	100	0.50	100	0.50	100	0.50	100	0.50				
$\frac{3}{4}$ in.	40	8.00			97	3.00	95	3.00	95	3.00				
$\frac{1}{2}$ in.	4	3.50	42	8.00							100	0.50	100	0.50
$\frac{3}{8}$ in.					17	6.00	37	8.00	37	8.00	90	5.00	100	1.00
No. 4	0	1.50			1	1.00					35	6.00	55	7.00
No. 8			1	2.00			5	2.00	2	2.00				
No. 16					0	1.00								
No. 100											0	1.00	0	1.00

Sieve Size	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ	$\overline{X}'$	σ
Aggregate Number	20		21		22		23		24	
$\frac{1}{2}$ in.										0.50
$\frac{3}{8}$ in.		0.50				0.50	100			2.00
No. 4		1.00		0.50		2.00	95	2.00		
No. 8		2.50		2.50						6.00
No. 16		6.00		6.00		6.00	60	6.00		
No. 30		4.00		4.00						6.00
No. 50		2.50		2.50		2.50				
No. 100		1.00		1.00		1.00	0	1.00		
No. 200										2.00

Note: $\overline{X}'$ = desired average, σ = standard deviation.

TABLE 4

TYPICAL MODEL SPECIFICATION
(Gradation Only)

61B6 Aggregates: Aggregate No. 3

Gradation: Each lot of aggregate shall be graded to conform to the requirements given hereafter. The size of a lot will be the quantity required for 1 mile of pavement 24 ft wide, or the quantity required for a project, whichever is smaller. Five random samples will be obtained from each lot as the aggregate is being used. The method of sampling will be established by the engineer.

Sieve Size	Percentage by Weight Passing (control limits)		
	Desired Average	Individual Results	Average of Five Results
½ in.	100	98.8 to 100	99.5 to 100
1 in.	95	85.7 to 100	90.8 to 100
¾ in.	40	21.4 to 58.6	31.7 to 48.3
½ in.	4	0.0 to 12.2	0.0 to 7.6
No. 4	0	0.0 to 3.5	0.0 to 1.6

When test results exceed these control limits, the lot of aggregate may be accepted subject to an adjustment in the delivered price that will be determined according to the following schedule:

Sieve Size	Percentage by Weight Passing									
	Individual Results					Average of Five Results				
1½ in.	98.8 100.0	98.6	98.4	98.1	97.6	99.5 100.0	99.4	99.3	99.2	99.0
1 in.	85.7 100.0	84.3	82.9	80.1	76.4	90.8 100.0	90.2	89.6	88.4	86.6
¾ in.	21.4 58.6	18.6 61.4	15.8 64.2	10.6 69.8	2.8 77.2	31.7 48.3	30.4 49.5	29.2 50.8	26.7 53.3	23.4 56.6
½ in.	0.0 12.2	13.4	14.6	17.1	20.4	0.0 7.6	8.1	8.7	9.8	11.2
No. 4	0.0 3.5	4.0	4.6	5.6	7.0	0.0 1.6	1.8	2.1	2.5	3.2
Percent Excess	0+	-15+	-30+	-60+	-100+	0+	-15+	-30+	-60+	-100+

Percentage for Payment	100	99.5	99	98	95	90*	100	99	98	95	90	80*

*The engineer will direct whether to adjust at this figure and leave in place or remove and replace.

The percentage for payment will be determined for each deviation in individual results and in the average of five results from the lot. For each sieve size only the smaller of the values thus obtained will be applied in adjusting the price. The adjusted price for the lot will be the delivered price of the aggregate multiplied in series by the percentage for payment for each sieve size.

TABLE 5

EXAMPLES OF PRICE ADJUSTMENTS

Sample No.	Sieve Size			
	1 in.	¾ in.	½ in.	No. 4
1. Test Results: Five random samples from lot No. 4; variations from control limits are underlined				
16	89.9	50.6	8.5	1.5
17	91.3	60.8	19.8	5.3
18	88.5	36.1	6.4	2.0
19	85.4	45.3	11.3	4.3
20	80.8	48.1	6.5	1.6
Average 5 results	87.2	48.2	10.5	2.9
Percentages for payment (from schedule in model for specification):				
Individual results	99.5 98.0	99.5	95	98 99
Average 5 results	90	100	90	90

Price adjustment: Assume delivered price $2.85 per ton
Price for payment = $2.85 × 0.90 × 0.995 × 0.90 × 0.90 = $2.07

Sample No.	1 in.	¾ in.	½ in.	No. 4
2. Test Results: Five random samples from lot No. 5; variations from control limits are underlined				
21A1	96.8	45.2	10.0	1.3
22A1	89.9	39.2	3.6	0.5
23A1	85.7	24.8	1.5	0.4
24A1	92.8	20.4	1.2	0.1
25A1	84.2	23.1	0.8	0.2
Average 5 results	89.9	30.5	3.4	0.5
Percentages for payment (from schedule in model for specification):				
Individual results	99	99.5	100	100
Average 5 results	98	99	100	100

Price adjustment: Assume delivered price $2.85 per ton
Price for payment = $2.85 × 0.98 × 0.99 = $2.76

for the lot of aggregate on a percentage basis that decreases as the magnitude of the deviation increases.

Each suggested specification is complete within itself except that the engineer or someone familiar with the theoretical considerations involved should select the location from which the random samples will be drawn. The guiding principle for this selection is that the samples will be drawn at a place as near as possible to where the aggregate is used in the work, and where the gradation of the aggregate is the same as it is when the aggregate is incorporated into the work. For concrete aggregate, sampling from a stopped loaded belt between the storage and the weighing bins is preferred. For surface treatment aggregate, sampling with special trays placed on the roadway is expected. For macadam or stabilizer aggregate, the sampling will be done at a randomly selected point after the aggregate has been spread and processed.

A typical model for a specification is given in Table 4. The method for determining the adjustment in price for actual test results is given in Table 5.

ACKNOWLEDGMENT

This paper is condensed from the final report on Research Project No. 520, conducted by the South Carolina State Highway Department in cooperation with the U.S. Department of Transportation, Federal Highway Administration, Bureau of Public Roads. The opinions, findings, and conclusions expressed in this report are the authors' and are not necessarily those of the Bureau of Public Roads or the South Carolina State Highway Department.

Statistical Study of the Compliance With Specification of Concrete Supplied for Highway Structures in the United Kingdom

D. H. MATHEWS, Road Research Laboratory, Great Britain; and
J. B. METCALF, Main Roads Department, Queensland, Australia

A study has been made of the statistical implications of a United Kingdom specification for strength of concrete for highway structures and the quality of materials supplied to and accepted under the specification to see in what way the specification might be improved. The distribution of results from strength tests is normal and the proportion of concrete defective at any specified strength may be estimated once the statistical parameters of the distribution have been evaluated. The proportion defective provides a useful measure of equality that can be used to study the effect of particular specification requirements on the quality of concrete supplied if the operating characteristic (OC) curve (relating the proportion defective and the probability of acceptance of that quality) is computed over a range of qualities.

The conclusions reached from examination of the OC curve are compared with the results of a statistical examination of the actual quality of concrete supplied to 186 jobs where 8,400 test results were obtained. Methods of improving specifications are briefly considered in the light of the risks to the producer of concrete and to the consumer, the object of any specification being to provide a fair apportionment of the risks of rejection of "good" material and acceptance of "poor" material between producer and consumer.

•SOME MEASURE of strength is used in most countries as an index of the quality of concrete. It is common practice to specify the strength required of concrete and to compare this with the actual strengths achieved as estimated by testing samples of the production. In the United Kingdom, for example, it has been the practice for many years to make test cubes from freshly mixed concrete and then, after curing for 28 days, to test these cubes in compression to provide information on the quality of the concrete before it was placed. The necessity of curing the test specimens means that the results are of little value in the control of the concrete production, which is best effected by other means, but such results are valuable to the purchaser or specifying authority who wishes to be assured that concrete of adequate quality has been produced.

The simplest approach to specifications is to state a "minimum" strength below which no test result may fall, and indeed many such specifications are still in use. However, there has been an increasing recognition that, because of the variable nature of the results on which acceptance decisions are based, this is an unsatisfactory and inefficient method of specification. As a result, there has been a movement in specifications toward the concept of a "characteristic" strength below which not more than a

Paper sponsored by Committee on Quality Assurance and Acceptance Procedures and presented at the 48th Annual Meeting.

fixed and predetermined proportion of the test strengths should fall. This recognizes that concrete production, sampling, specimen making and curing, and testing are all variable procedures and will contribute to the overall variability of the final test results on which decisions are based.

This variability can be taken into account in specifications relatively simply because measurements of concrete strength can be treated as being normally distributed (1, 2). Occasionally it is possible to show that a particular set of results is not distributed normally, but such distributions probably arise from nonrandom variations in the production. Where it is known, for example, that changes in aggregate or in the design composition of the mixture have taken place, these can be taken into account during the statistical treatment, but some variations may occur unknown to or unrecorded by the producer. However, if the assumption of normality is not strained too far by treating results of a very long period of production as homogeneous or by placing undue emphasis on the frequency of occurrence of results in the extreme tails of the distribution, a satisfactory practical basis for statistical treatment is available.

A normal or Gaussian distribution may be defined completely in terms of a mean (μ) and a standard deviation (σ). It follows that a specified or characteristic strength of concrete may be defined as

$$L = \mu - k\sigma$$

where k is a constant defining the proportion of the overall distribution falling below L, i.e., the proportion defective (p). In the practical writing of a specification, a value of σ is assumed and μ is fixed by the design requirements. Thus, for a given "failure rate" denoted by k, a specified strength, L, may be calculated. However, when it comes to the judgment of compliance with this specified strength, decisions have to be based on estimates, m and s, of the true mean and of the standard deviation achieved. These estimates are derived from sampling the overall population of possible strengths and it is here that the uncertainty arises. Decisions are often required on relatively small numbers of results where the values of m and s may not be good estimates of μ and σ.

There are two general methods of approaching the judgment of compliance with specification; these methods have been discussed in relation to the composition of bituminous mixtures by Mathews and Hardman (3). In the case of strength of concrete, test results may be classified (a) by attributes, i.e., employing the numbers of test results that fall short of the specified strength; or (b) by variables, i.e., employing the magnitudes of the test results to estimate the true mean strength and sometimes the variability to compare with a specification of these properties. The second method is normally used on small groups of measurements where the mean and the standard deviation (or sometimes the range) are computed and decisions are based on a comparison of $m_n - k's_n$ with the specified strength. This approach is efficient in the use of data but requires some computational effort. In general, the value of the constant k′ should be different from the value of k used in design that applies only to the infinite distribution; the value required, k′, is that applying to the actual group size, n.

RISK IN SPECIFICATION

The aim of any specification is to state clearly the minimum quality of material or work that is required. The method adopted for judging compliance should ensure that work of the specified minimum quality or better is accepted and work of a worse standard is rejected. Unfortunately, when decisions are based on sampling from the overall population of possible results, a clear-cut decision free from all risk of error is impossible. It is possible to be only reasonably sure.

The decisions on compliance with specification must be fair to the producer, who requires assurance that when he produces work of the specified quality it is likely to be accepted, and to the consumer, who requires reasonable assurance that his specification is being met and that he is getting value for money. In neither case can the assurance be absolute, and therefore in any scheme based on sample measurements there are two risks to be recognized and assessed. There is a "producer's risk" that "good" material may be rejected, which forms part of his overall commercial risk, and there is a "consumer's risk" that "bad" material may be accepted. The most important factor in determining these risks is the rate of sampling. Where the tests are few, the risks are high, especially for the consumer if in the design of the acceptance scheme the view is taken that the producer's risk should be fixed at a low level to minimize the risk of rejection of suitable material. This leads to the conclusion that the design of the acceptance scheme and the accompanying rate of sampling by the consumer should pay considerable attention to the criticality of the design requirements. At present little or no attention is paid to this point in specification.

The risks involved can be seen more clearly by reference to an example. If it is specified that concrete be supplied with no more than 5 percent of the total production less than the specified strength, L, and then if decisions are taken on the basis that a single sample is tested and the material is accepted if the result is not less than the specified strength, there is a risk to the producer that on the average the material of the minimum specified quality will be rejected once in twenty times. On the other hand, if a production with a mean strength equal to the specified strength is offered (that is, material which is 50 percent defective), then there is an even chance that the single test will fall above the mean and the production will be accepted. In these circumstances, therefore, the consumer has a 50 percent risk of accepting a production that is 50 percent defective. In this example there is a linear relationship between the true proportion that is defective and the chance of acceptance of that quality. This is shown by the straight line in Figure 1; it is the simplest example of an operating characteristic (OC) curve. Such OC curves, which define the relationship between the defective level and probability of acceptance, can be derived for any type of acceptance scheme.

The ideal form of the OC curve for an acceptance scheme would be such that all qualities of production better than the specified minimum would be accepted, and all worse qualities would be rejected all the time. This is impossible in practice and it is necessary to have a relationship where there is some producer's risk and a consumer sumer's risk, which may be very large indeed if decisions are based on small numbers of test results.

OPERATING CHARACTERISTIC CURVE FOR STRENGTH OF CONCRETE SPECIFICATION

The probability of acceptance with an attributes scheme is calculated relatively easily using the binomial expansion of $(p + q)^n$ where p is the true average proportion defective, $(p + q) = 1$, and n is the number of samples on which a decision is based. These calculations are described and typical results tabulated in most of the standard statistical texts. The calculations of the probability of acceptance for a variables scheme of the type "accept if $m_n - k' s_n \geq L$", where m_n and s_n are the mean and standard deviation of a set of n results, presents rather more difficulty because the proportion defective is determined by two variables, m_n and s_n, one of which, m_n, is distributed normally while the other, s_n, has a skew distribution, so that the mean estimate does not coincide with the population standard deviation, σ. This type of problem is also described in statistical texts, and Resnikoff and Lieberman (4) give tables that make possible the necessary estimates of probability of acceptance over a range of true qualities.

The evaluation of the OC curve for the assessment of an acceptance scheme may

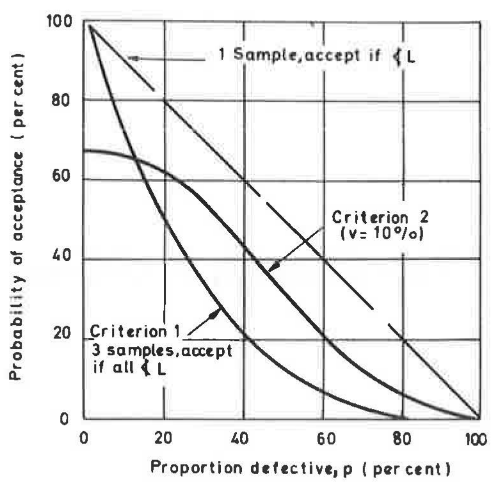

Figure 1. Examples of some operating characteristic (OC) curves.

best be shown by reference to an actual specification. A specification that has been widely used in the United Kingdom for structural concrete takes the following form:

> Consider the results of strength tests three at a time: if none of the three is less than the specified strength (L), accept; if any result is less than L, calculate the mean (m_3) and the range (r_3) of the three results and then accept if both m_3 is not less than L and r_3 is not greater than one fifth of m_3.

Clearly this specification was intended to allow some test results to fall below the specified strength and in such circumstances still classify the concrete as acceptable if the mean strength is at least equal to that specified and the concrete is not excessively variable. The OC curve for the first criterion is shown in Figure 1, which also shows the OC curve for the second (double) criterion. This second criterion operates only on concrete that has not been accepted by the first criterion, and thus the overall probability of acceptance for the two criteria taken together is given by

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

where P_1 is the probability of acceptance by the first criterion and P_2 the probability of acceptance of the same quality by the second criterion. The individual and overall probabilities are given in Table 1 for a concrete with a coefficient of variation of 10 percent. Figure 2 shows the effect of the range requirement for concretes of the same true proportion defective produced under two different standards of control corresponding to coefficients of variation of 10 and 15 percent.

It will be seen from the OC curves that an output some 30 percent defective to the specified strength will be accepted 60 percent of the time in spite of the relatively complex acceptance scheme. However, the producer needs to maintain the level of defectives below about 4 percent to have at least a 95 percent chance of the work being accepted. Thus, the net result of the application of the two criteria is to produce an OC curve that, over most of the range of possible qualities, provides little better protection to the consumer than the single test criterion shown by the straight line in Figure 1. This seems to be a very poor return for the complexity of the scheme.

DEFINITION OF ACCEPTABLE QUALITY

It is clear that no single parameter will adequately define the acceptable strength of concrete; the process of judging compliance on test samples must mean that a range of strengths will be accepted and therefore in writing a specification it is necassary to

TABLE 1

INDIVIDUAL AND COMBINED PROBABILITIES OF ACCEPTANCE
(Coefficient of Variation 10 percent)

Percent Defective	Probability of Acceptance		Combined Criteria
	Criterion 1	Criterion 2	
P	P_1	P_2	$P_1 + (1 - P_1) P_2$
0.0	100	67	100
1.0	97	67	99
2.5	93	67	97
5.0	86	67	95
10.0	73	66	91
20.0	51	62	81
25.0	43	58	76
35.0	28	49	63
50.0	13	33	42
60.0	7	22	27
70.0	3	13	16
80.0	1	5	6
90.0	-	1	-

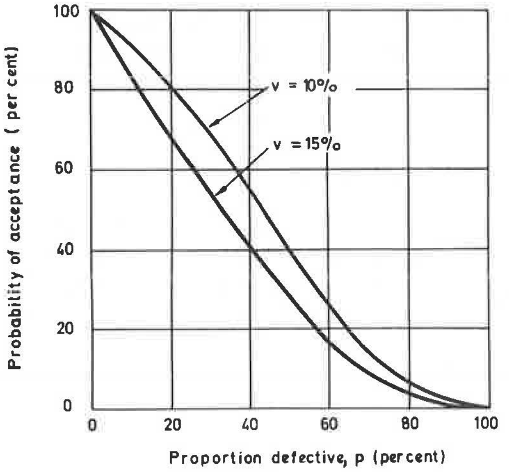

Figure 2. Combined OC curves at two levels of variability for criterion 1 and criterion 2 taken together.

define this range of strength. Ideally, this strength distribution should be so related to the distribution of stresses to which the concrete might be subjected in service that the risk of a low strength coinciding with a high stress would be negligible, but this is not yet possible. The alternatives therefore are either to set some arbitrary standard, as at present, or to examine what is economically possible in current practice and to accept this as a valid estimate of desirable quality. A specification may then be drafted to ensure acceptance of no worse quality than this, and to exert pressure for a change in the quality level where justified.

The levels of strength attained in current practice in the United Kingdom were therefore investigated. The strength-test results for concrete supplied for a substantial number of highway structures built to the specification discussed previously were examined and the quality expressed in terms of the proportion defective estimated from the properties of a normal distribution.

Two assumptions have been made. One assumption is that the proportion defective calculated from the test results is a reliable estimate for any particular lot of concrete. The errors associated with this assumption will be small except (a) where inhomogeneity occurs in the data (this has been detected in only very few cases where extremely long runs of production have been examined; in such cases, assignable causes of variation clearly exist but insufficient evidence is available to eliminate their effect on the population parameters), and (b) where the samples have not been taken at random (for example, where the engineer samples specially from what he believes to be poor concrete; in the present case, the error from this source was probably small). Two further minor sources of error occur when (c) concrete is rejected for reasons of workability (this tends to curtail the extremes of the distribution), and (d) low test results are discarded (because it is usual to find low results specially noted in site records, this is not likely to be a source of serious error).

The second assumption is that, because the concrete has been accepted, the distribution of strengths observed adequately describes the desirable level of quality. This introduces two sources of possible substantial error because (a) the data show that on a number of occasions where relatively low strength was specified, the concrete supplied had high strength (presumably it was cheaper or more convenient on site to use a supply of concrete already available than to produce a relatively small amount of the specified grade); and (b) rejection of concrete solely on grounds of inadequate strength appears to be extremely rare (the data analyzed showed no instances of rejection although it was clear that in some cases the level of strength was subsequently increased by a distinct amount when low test results had been obtained).

In spite of the reservations arising from the foregoing assumptions, it is believed that the distribution of the level of defectives derived from the survey of current practice gives a valid estimate of the quality that is currently regarded as acceptable by engineers for highway structures. The distribution of quality obtained is shown in Figure 3. Comparison of Figures 2 and 3 indicates that the actual quality obtained in practice was somewhat better than the quality that could have been accepted under the specification; the average level of defectives is about 4 percent, and 90 percent of the work was less than 10 percent defective. The highest proportion of defectives was 30 percent. It has been found from an examination of the data from the 186 jobs, that neither the level of specified strength nor the method of production had any significant effect on the distribution of quality.

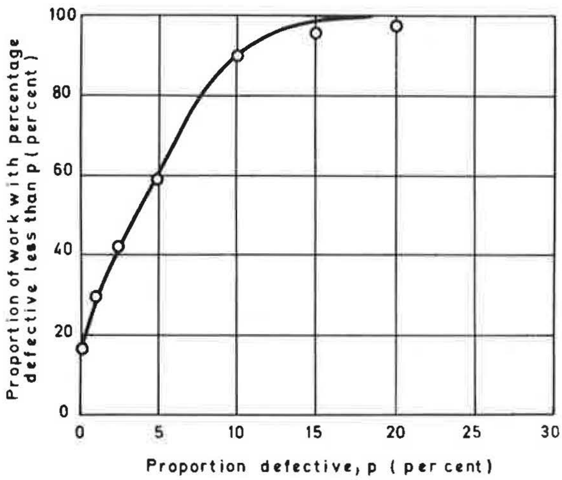

Figure 3. Distribution of proportion defective for 186 jobs (8,403 tests).

Two other measures of the strength distribution can be derived from the data: (a) the distribution of the ratio of the mean to the specified strength (m/L), and (b) the distribution of the variability, expressed as the coefficient of variation (σ/μ). These are shown in Figures 4 and 5. It is necessary to make a distinction between low- and high-strength concrete because there is a practical limit to the strength attainable, and therefore for the high-strength concrete the ratio of mean to specified strength is smaller on the average. Furthermore, to achieve regular compliance with a high-strength criterion, it is necessary to have good control, and therefore the coefficient of variation tends to be small. However, the curves for the lower strength concretes indicate that the high average quality commented on earlier was achieved in the main by working with an average strength well above that strictly necessary rather than by applying very strict site control.

DISCUSSION AND CONCLUSIONS

It is one thing to specify a desired strength of concrete; it is another to obtain reasonable assurance for the consumer that the specified strength is attained and at the same time provide a reasonable assurance to the producer that good-quality material is not rejected. In the example cited in this paper, there is little apparent connection between the quality that is found in practice and the quality that the specification might be expected to give.

With concrete in highway structures in Great Britain, the quality is such that, on an average job, 4 percent of the concrete is likely to be less than the specified strength; one would expect that if producers just met the specification, the proportion defective should be somewhat higher. This situation would seem to result from two factors: one,

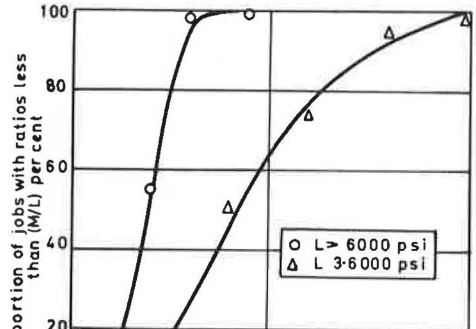

Figure 4. Distribution of ratio of mean to specified strength (m/L).

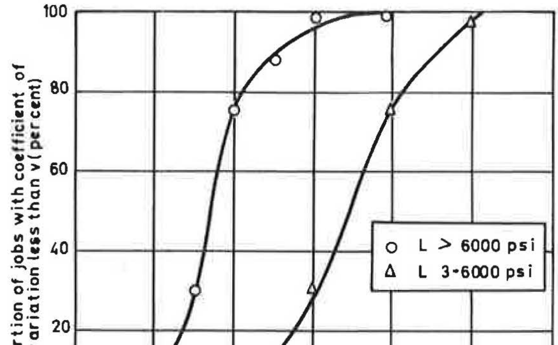

Figure 5. Distribution of coefficient of variation (V=100 s/M).

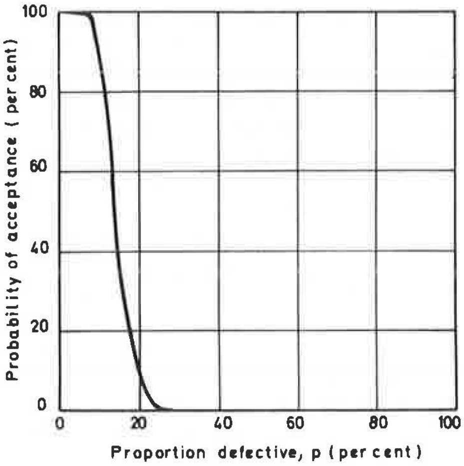

Figure 6. OC curve for criterion "accept if $L \leq m_{30} - 2/3\,(\bar{r}_3)$".

the exercise of judgment by engineers (outside the specification requirements), and the other, and probably more important, the desire of producers to limit their risks severely because of the serious economic consequences of rejection. On a consideration purely of production costs and selling prices, it might be concluded that the break-even point should be reached at about 80 percent probability of acceptance, which in the present example means a defective level of about 15 percent. However, the penalties associated with rejection are likely in general to be considerably greater than the production cost of the rejected material, and hence the contractor is unwilling to operate at a level of quality that gives him any significant risk. This suggests that, if the level of quality found in the survey is what is really required, the specification should be recast to provide such a quality and not to rely on intangible factors.

There is considerable scope for the examination of the risks to the consumer implicit in specification requirements because these are frequently not appreciated. Even the quite complex requirements discussed in this paper do not provide adequate protection in a really critical situation. Indeed, no method of judgment based on only a very small number of tests can ever provide a really high degree of assurance. The testing requirements have always to be balanced against the economic consequences of undetected poor-quality material. The risks can be limited by the adoption of efficient criteria of judgment, such as the comparison of means of small groups of results—say, three or four at a time—with a limit set in relation to the specified strength, group size, and some fixed level of producer's risk. Even greater efficiency can be obtained if it is possible to use a "variables" scheme rather than an "attributes" scheme. For example, a criterion for ten successive groups of three test results is of the form

$$L \leq m_{30} - \frac{2}{3}\,\bar{r}_3$$

where m_{30} is the mean of the 30 results and $\bar{r}_3$ is the average range of the ranges in the ten groups of three. This would be an efficient means of maintaining a distribution of quality similar to that described in Figure 3. The OC curve for this criterion is shown in Figure 6.

The most important factor too often neglected in the implementation of compliance schemes is that the decisions must be consistent and consistently enforced. It is only in this way that the risks will be fairly apportioned between the producer and the consumer. If the decisions are distorted in any way, such as by accepting, albeit reluctantly, material indicated as unsatisfactory or by rejecting material on insufficient grounds, then the inevitable consequence will be to distort the risks.

ACKNOWLEDGMENTS

The work described in this paper forms part of a larger study of the efficiency of specifications, control, and compliance with specification that is in progress at the Road Research Laboratory. This work was carried out during the assignment of one of the authors to the Road Research Laboratory from the Main Roads Department of Queensland and the Australian Road Research Board. The authors wish to acknowledge the assistance given by Mr. R. Hardman in the development of some of the statistical arguments and that provided by Miss P. Damji in computation. Contributed by permission

REFERENCES

1. A Guide to Specifying Concrete. Institution of Civil Engineers, London, 1967.
2. Erntroy, H. C. The Variations of Works Test Cubes. Cement and Concrete Association, Research Rept. No. 10, London, 1960.
3. Mathews, D. H., and Hardman, R. Proc. Nat. Conf. on Statistical Quality Control Methodology in Highway and Airfield Construction, Univ. of Virginia, Charlottesville, 1966.
4. Resnikoff, G. J., and Lieberman, G. J. Tables of the Non-Central t Distribution. Stanford Univ. Press, Stanford, California, 1957.

In-Service Degradation of Base Course Aggregates

KARL H. DUNN, Research Engineer, Wisconsin Department of Transportation

A limited field study was conducted to determine the amount of degradation that may occur in untreated aggregate base courses as a result of manipulation, compaction, and service exposure. The principal base aggregate types available in Wisconsin—hard and soft crushed dolomite and dolomitic and igneous gravels—were included in the study.

Assessment of degradation was based on changes in the average gradations of multiple samplings of test sections before manipulation (as-produced), after compaction, and after two and five years of service exposure.

The test data indicate that the greatest amount of degradation of the aggregates occurred during manipulation and compaction, especially in the base courses placed in single lifts. The increase in the amount of clay-size particles was small (under 1.5 percent), and no change from the original "nonplastic" condition was experienced. No relationship was found between the magnitude of degradation and the type of physical properties of the aggregates.

•NUMEROUS investigations of flexible pavement failures in Wisconsin during the period from 1957 to 1960 indicated that failure was partially caused by lack of adequate base support. Results of tests of the in-place base materials showed many instances of excessive fines (material passing the No. 200 sieve) or excessive plasticity. Because tests on these same materials during construction did not give evidence of these excessive values, it was believed that degradation or disintegration of the materials had occurred after having been placed in service.

An investigation of possible in-service degradation was therefore proposed and accepted for inclusion in Wisconsin's HPR program. The materials research unit initiated the investigation during the 1962 construction season. The projects were selected to include the principal types of aggregates available in Wisconsin. The investigation included sampling of base course aggregate immediately after spreading, after compaction, and after two and five years of service. The results of tests on samples obtained at these various intervals of service are tabulated and discussed in this report.

REVIEW OF PREVIOUS WORK

Available literature on the subject of aggregate degradation in relation to highway pavement material use was limited at the time this study was initiated. The information that was available concerned laboratory studies of the degradation process (1) or general literature reviews of the subject preliminary to proposing further studies (2). There was almost no information published on in-service degradation of aggregates except in some regions where abundant local materials had shown excessive degradation (3).

The majority of degradation information was concerned with surface aggregates rather than base or subbase materials. This was probably because degradation was generally more prevalent on the surface where the abrasive forces act. Also, the contact surface loads induced higher stresses than those at lower depths in the pavement

Paper sponsored by Committee on Mineral Aggregates and presented at the 48th Annual Meeting.

structure. Thus, most studies had been concerned with bituminous mixtures or surface treatment mixtures.

The availability of literature was further restricted in that almost no information was available concerning degradation during construction, even though the fact that aggregates did degrade had been recognized for many years. Although most highway agencies had noted some degradation of aggregates, investigations had been made only by those agencies experiencing severe problems with regard to degradation. The term "degradation" was generally associated with aggregate breakdown into finer sizes. Specific interpretations of how this action occurred varied. Certain areas (3, 4) had observed breakdown caused by the action of water or what might be termed "hydraulic action". In those areas, plastic fines were produced as a result of basalt aggregates degrading to form montmorillonite clay minerals. More often, aggregate degradation was associated with "mechanical actions" such as those produced by abrasion and impact resulting from handling, compaction, and service. These latter actions were the types that were believed to be the primary causes of degradation of certain Wisconsin aggregates.

Measurements

All aggregate materials are theoretically capable of degrading to a maximum density, provided that the acting forces and other contributing factors are of the required intensity. Moavenzadeh and Goetz (1) concluded that the pattern of degradation was essentially a constant and could be measured or observed by gradation curves and sieve analysis data.

The pattern of degradation is distinct from the magnitude of degradation. Fracture or degradation patterns are important in mining operations, whereas magnitude of degradation is more applicable to engineering uses. An exact measure of the magnitude of degradation would be very difficult and probably not practical to achieve. However, Moavenzadeh and Goetz (1) found that the percent increase in surface area was a satisfactory measure of degradation, provided that realistic surface area factors were used. The amount of P-200 material (material passing the No. 200 sieve), as measured by a washed gradation, also was believed to be a simple and practical measure.

If the possibility of plastic property changes exists in fines as a result of degradation, it may be necessary to combine several measures to present the overall results. A hydrometer analysis would indicate whether the P-200 fines were increasing in the silt or clay sizes.

Affecting Factors

Pertinent factors that were generally cited in the literature as controlling degradation were type of aggregate, maximum size and gradation of particles, aggregate shape, compactive effort, subbase and/or subgrade influence, water, time, weathering, layer thickness, and orientation of particles (2). In the case of base course aggregates, another factor would be the type of surface. The relative effects of these factors would be difficult to evaluate in the laboratory and would require extensive and time-consuming field studies. This investigation was therefore limited to determining the amount of degradation of a base course aggregate, and studying the effect of origin and type of base course material on the amount of degradation and the change in plasticity resulting from handling, compacting, and service.

PURPOSE AND SCOPE

The basic objective of this study was to determine the amount of degradation, if any, of base course aggregates that may take place during construction and during subsequent service life in Wisconsin. The effect of origin and type of base course aggregate on the amount of degradation was also to be evaluated.

The selection of projects for sampling was limited to flexible pavements consisting of hot-plant bituminous pavements placed on untreated crushed stone or on gravel base courses constructed during the 1962 construction season. At the time the study was

TABLE 1

DESCRIPTION AND LOCATION OF TEST SECTIONS

Test Section	Type of Highway	Location	Type of Base Course Material and Number of Lifts
1	County Trunk Highway (CTH-H)	Racine County, southeastern Wisconsin	Dolomitic gravel, two lifts
2	State Trunk Highway (STH-31)	Racine County, southeastern Wisconsin	Crushed dolomite, Niagara formation, three lifts
3	County Trunk Highway (CTH-G)	Juneau County, central Wisconsin	Crushed dolomite, Prairie du Chien formation, single lift
4	State Trunk Highway (STH-48)	Barron County, northwestern Wisconsin	Igneous gravel, two lifts
5	County Trunk Highway (CTH-Y)	Clark County, central Wisconsin	Igneous gravel, single lift

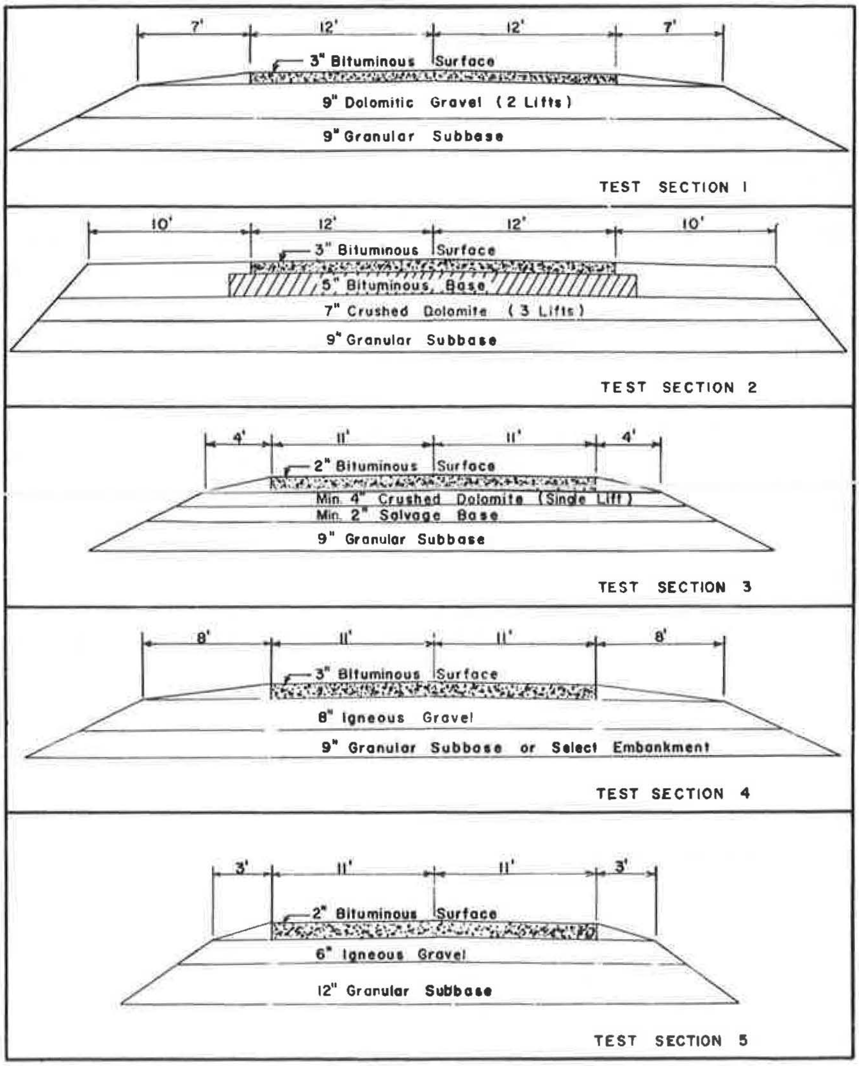

Figure 1. Typical test sections.

initiated, the design criteria stipulated that base courses of flexible pavements be stabilized if the anticipated loadings would exceed 50 heavy commercial vehicles per day. Also, the sodium sulfate soundness test was being inaugurated as a requirement for base course aggregates. The adopted specification limited the loss to a maximum of 18 percent after five cycles of the test. Many aggregates from sources known to have previously produced aggregate bases with poor service performance were not meeting these requirements, so these aggregates could not be included in the study. As a result of these limitations, it was possible to include only five projects in the study. However, the aggregates used for these selected projects did include the principal types of base course aggregates available in Wisconsin. The locations and general descriptions of these projects are given in Table 1. A typical cross section for each of the five test sections is shown in Figure 1.

Variability can be expected between gradations of individual samples obtained within close proximity. This variability is caused by variances such as chance, sampling, testing, and the inherent variability of the material itself. Consequently, numerous samples were obtained from a given unit of roadway and the average gradation of these samplings was considered the representative gradation of material within the unit. Assessment of degradation was based on changes in these average gradations after selected intervals of time. Material sampling and testing variances were therefore considered to be essentially constant, and changes in average gradations were indicative of actual degradation.

MATERIALS

The types of aggregates included in this study were dolomitic gravel, crushed dolomites from the Niagara and Prairie du Chien formations, and igneous gravels. Lithological analyses of these materials are given in Table 2.

PROCEDURES

Field Sampling

Sampling of the base course aggregates was accomplished within a 24- by 1000-ft test section area on each project. An attempt was made to locate this section near the project midpoint, between the source of aggregate and the most distant point of haul. The factor of amount and type of traffic during and after construction was therefore considered average for the individual project, but different from project to project. The restriction of sampling within a limited area also served to reduce possible variations caused by differences in exposure, weathering, and methods used for placing and compacting the material.

TABLE 2

LITHOLOGIC DESCRIPTIONS

Test Section	Type of Aggregate	Lift	Rock Types, Percentages by Weight										
			Igneous					Sedimentary				Metamorphic	
			Basalt	Diabase	Granite	Poryphyry	Rhyolite	Chert	Siltstone	Sandstone	Dolomite	Schist	Quartzite[c]
1	Dolimitic	Top	—	14[a]		—	—	2	1	—	81	—	2
	gravel	Bottom	—	16[a]		—	—	2	2	—	80	—	—
2	Crushed	Top	—	—	—	—	—	—	—	—	100	—	—
	dolomite,	Middle	—	—	—	—	—	—	—	—	100	—	—
	Niagara	Bottom	—	—	—	—	—	—	—	—	100	—	—
	formation	Bottom	—	—	—	—	—	10	—	—	90	—	—
3	Dolomite, Prairie du Chien formation	Single	—	—	—	—	—	Tr[b]	—	—	99+	—	—
4	Igneous	Top	41	4	4	10	—	11[d]	—	11	—	—	19
	gravel	Bottom	26	16	9	—	—	3[d]	—	—	—	3	36[e]
5	Igneous	Single[g]	40	—	20[f]	—	20[d]	—	—	2	—	3	15
	gravel		31	13	19	10	—	5	—	2	—	—	20

[a]Indicates mixture of basic and acidic igneous rocks.
[b]Indicates trace, not over 1 percent.
[c]Includes quartz.
[d]Includes jasper.
[e]May include hard sandstone.
[f]Includes 5 percent weathered particles.
[g]Replicate samples.

A grid system of 3- by 5-ft rectangles was established within each test section area, resulting in a total of 1600 incremental areas from which sample locations could be randomly selected. One sample was obtained from each selected incremental area.

During construction, 30 samples were obtained both before and after compaction at each project. This number of samples was considered sufficient to provide a reliable estimate of the true average gradation. Each sample weighed approximately 25 lb.

Wisconsin specifications state, "The work shall, in general, proceed from the point on the project nearest the source of supply of the aggregate in order that the hauling equipment will travel over the previously placed material, and the hauling equipment shall be routed as uniformly as possible over all portions of the previously constructed courses or layers of the base course." In a project requiring two or more layers of base course, the material in each layer could be different within any specific area because of differences in traffic exposure and possible changes of source of aggregate within the pit or quarry during the time lapse between placement of each layer. Consequently, each layer (lift) was sampled as a separate entity.

Sampling of the loose material prior to compaction consisted of locating the proper grid sampling point by random selection, and obtaining approximately 25 lb of material by use of a square, flat-bottomed shovel. To control variation caused by sampling, care was taken that no material was lost from the shovel during sampling. Generally, these samples were obtained immediately after deposition from the trucks, but in some instances a motor grader had made an initial pass to spread the material.

Samples of the compacted material were generally obtained within a day after completion of compaction; however, weekends or inclement weather occasionally delayed sampling for several days. Traffic during these periods of delay was relatively light and not regarded as producing a significant increase in degradation beyond that produced by compaction.

An area of approximately 8 by 8 in. of the compacted material was removed as an individual sample. The material within this area was loosened by a pick or a long chisel formed from a $^3/_8$-in. reinforcing bar. Use of the pick and chisel were kept to a minimum to avoid possible degradation. A square-end trowel was used to remove the loosened material from the sample area. Material was removed to a depth about 1 in. less than the thickness of a compacted lift. This volume of compacted material yielded the desired sample weight.

The void remaining after sampling had vertical sidewalls and a flat bottom. The objective of this sampling procedure was to be sure that all material was removed, including all fines.

At least three 50-lb samples were obtained at each pit or quarry for use in determining the physical properties of the aggregates. These samples were obtained at the same time as the material was being placed within a respective test site area. This sampling was repeated for each individual layer of base course aggregate placed on a project.

Preliminary review of test results for the original sampling indicated that equivalent reliability of results could be obtained with fewer but larger samples. Because in-service sampling required removal of portions of the pavement surface, it was desirable to reduce the number of such samples. Consequently, only 12 samples of 75 lb each were obtained within a given test area after two and five years of service. Sample areas were randomly selected essentially in the same manner as previously described, except that no individual grid area previously sampled was resampled, and the selected area corresponded to either the outside or inside wheelpaths. Sampling was limited to the wheelpaths because degradation was expected to be prevalent within these areas of the bases.

The general sequence of sampling after two and five years was as follows:

1. Remove surface with air hammer;
2. Remove $^1/_2$ to 1 in. of top of base material leaving a level surface (to ensure that no asphaltic material was included in the sample because this material would hinder the sieving operations);
3. Loosen material within an approximate 2- to 3-ft square area and 2 to 4 in. deep for sample of top lift (keeping use of the pick or chisel to a minimum to avoid excessive disturbance of the material);

4. Remove all material from the sampling area and place in a sample bag; and
5. If two lifts were to be sampled, discard about 1 to 2 in. of the material to be sure that the second sample would be from the proper thickness of compacted material.

Laboratory Testing

A special technique of conducting wash gradings was used to ensure that all fines in the samples were accounted for within the proper sieve size fraction. This sequence involved several washings of the material and repeating certain steps if the results for two samples were not in close agreement.

At least three hydrometer analyses were conducted for each series of wash gradations used to establish the average gradation of an individual lift. Results of these individual hydrometer analyses were averaged to determine the percentages of silt and clay present in each respective lift. Atterberg limits testing was also conducted to determine the plasticity of the fines. These hydrometer and plasticity tests were conducted in accordance with applicable AASHO standard test procedures.

METHODS OF ANALYSIS

Average gradation results were separated and arranged into particle-size groupings to obtain comparisons of the changes in the distributions of the particle sizes produced by construction and service. These comparisons allowed an evaluation of the patterns of degradation for the aggregates.

It was pointed out previously that a comparison of gradations may show a pattern of degradation, but an evaluation of the magnitude of degradation requires some type of numerical measure. Two numerical measures were adopted for this study as follows: (a) percent of material passing the No. 200 sieve (referred to as P-200), and (b) surface area.

The surface area values used in this study involved the use of special conversion factors that were based on the assumptions that all material passing the No. 4 sieve was spherical and that the material retained was one-third cubes and two-thirds parallelepipeds with sides of 1:2:4 proportions, as described in a report by Moavenzadeh and Goetz (1).

Values used for computing surface areas were obtained from average particle-size accumulation curves because only selected sieve sizes were used to establish the average gradations. The P-200 material values were taken directly from the average gradations. The hydrometer and plasticity test results were used to determine if changes in the composition of fines took place and if these changes produced plasticity.

TEST RESULTS

Results of tests conducted on samples of material obtained at the site of production are summarized in Table 3. These data provide for comparison of the physical properties of the aggregates.

TABLE 3
SUMMARY OF PHYSICAL PROPERTIES[a]

Property	Test Section								
	1		2			3	4		5
	Top	Bottom	Top	Center	Bottom	Single	Top	Bottom	Single
Bulk specific gravity[b]	2.68	2.68	2.69	2.70	2.68	2.64	2.72	2.70	2.71[f]
Absorption[b]	1.57	1.67	1.04	1.00	1.18	2.08	1.17	1.17	1.50[f]
Percent wear, 100 revolutions[c]	12	6	7[f]	8[f]	7[f]	9[f]	6[f]	5[f]	6[f]
Percent wear, 500 revolutions[c]	29[f]	29	27[f]	29[f]	30[f]	40[f]	23[f]	20[f]	23[f]
Soundness[d]	7.1[f]	4.6	0.7[f]	1.2[f]	5.0[f]	6.6[f]	3.2[f]	2.0[f]	5.1[g]
Plasticity index[e]	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.

[a]Samples obtained either at plant, pit, or stockpile.
[b]Specific gravity and absorption of coarse aggregate, AASHO T 85.
[c]Los Angeles abrasion of coarse aggregate, AASHO T 96.
[d]Soundness of aggregate, AASHO T 104-57.
[e]AASHO T 89, T 90, T 91.
[f]Average of two tests.
[g]Average of four tests.

TABLE 4

AVERAGE GRADATIONS, PERCENT BETWEEN SIEVES

Sieve Size	Test Section 1—Dolomitic Gravel							
	Sampled During Construction[a]		2 Years[b]	5 Years[b]	Sampled During Construction[a]		2 Years[b]	5 Years[b]
	Original	Compacted			Original	Compacted		
	Top Lift				Bottom Lift			
1 in. to 3/4 in.	5	5	4	5	5	4	4	4
3/4 in. to 1/2 in.	15	16	14	14	15	15	13	13
1/2 in. to 3/8 in.	9	8	9	8	9	8	9	8
3/8 in. to No. 4	14	15	14	15	15	15	14	15
No. 4 to No. 10	13	12	13	12	13	13	14	13
No. 10 to No. 40	21	22	21	21	20	20	21	21
No. 40 to No. 200	16.1	14.4	16.8	17.0	16.9	17.5	17.9	18.4
No. 200	6.9	7.6	8.2	8.0	6.1	7.5	7.1	7.6

Sieve Size	Test Section 2—Crushed Dolomite (Niagara)					
	Sampled During Construction[a]					
	Original	Compacted	Original	Compacted	Original	Compacted
	Top Lift		Center Lift		Bottom Lift	
1 in. to 3/4 in.	5	4	5	4	6	5
3/4 in. to 1/2 in.	21	17	20	18	24	22
1/2 in. to 3/8 in.	15	13	13	12	13	13
3/8 in. to No. 4	23	22	20	19	18	19
No. 4 to No. 10	12	14	12	14	12	13
No. 10 to No. 40	10	13	12	14	11	12
No. 40 to No. 200	4.5	5.4	6.4	6.3	6.1	6.2
No. 200	9.5	11.6	11.6	12.7	9.9	9.8

Sieve Size	Test Section 3—Dolomite (Prarie du Chien)				Test Section 5—Igneous Gravel			
	Sampled During Construction[a]		2 Years[b]	5 Years[b]	Sampled During Construction[a]		2 Years[b]	5 Years[b]
	Original	Compacted			Original	Compacted		
	Single Lift				Single Lift			
1 in. to 3/4 in.	6	4	4	4	5	3	4	4
3/4 in. to 1/2 in.	23	19	16	18	16	13	14	16
1/2 in. to 3/8 in.	14	12	12	12	10	9	9	9
3/8 in. to No. 4	21	21	21	22	14	13	14	14
No. 4 to No. 10	12	13	14	14	10	11	10	10
No. 10 to No. 40	8	9	11	10	25	27	25	24
No. 40 to No. 200	7.9	10.3	10.8	9.3	14.3	15.7	15.2	14.8
No. 200	8.1	11.7	11.2	10.7	5.7	8.3	7.8	8.2

Sieve Size	Test Section 4—Igneous Gravel							
	Sampled During Construction[a]		2 Years[b]	5 Years[b]	Sampled During Construction[a]		2 Years[b]	5 Years[b]
	Original	Compacted			Original	Compacted		
	Top Lift				Bottom Lift			
1 in. to 3/4 in.	7	8	7	6	6	6	6	7
3/4 in. to 1/2 in.	18	19	18	17	17	18	17	17
1/2 in. to 3/8 in.	10	9	11	10	9	9	10	9
3/8 in. to No. 4	15	16	15	15	14	15	14	15
No. 4 to No. 10	12	11	11	10	11	10	11	10
No. 10 to No. 40	21	20	19	22	22	22	21	21
No. 40 to No. 200	10.8	10.3	11.5	12.0	14.2	13.2	13.3	13.5
No. 200	6.2	6.7	7.5	8.0	6.8	6.8	7.7	7.5

[a]Average of 30 samples, 25 lb each: Original, sampled before compaction; compacted, sampled after compaction.
[b]Average of 12 samples, 75 lb each: Sampled after 2 and 5 years of service.

The thicknesses of bituminous concrete placed on the base courses varied from 2 to 3 in. for test sections 1, 3, 4, and 5. An additional 5 in. of bituminous base course was placed in test section 2, so that the total thickness of cover over this base course was 8 in. It was considered impractical to remove 8 in. of pavement to sample this base course; consequently, no sampling of test section 2 was attempted after placement of the bituminous base and surface courses. Evaluation of degradation of the crushed dolomite, Niagara formation, was therefore limited to that produced by compaction during construction.

Average gradations for each lift of each test section are presented in Table 4 on a percent between-sieves basis, which provides for interpretation of changes in gradation.

TABLE 5

PARTICLE SIZE DISTRIBUTION, PERCENT OF TOTAL MATERIAL

Particle Size and Description	Test Section 1—Dolomitic Gravel							
	Sampled During Construction[a]		2 Years[b]	5 Years[b]	Sampled During Construction[a]		2 Years[b]	5 Years[b]
	Original	Compacted			Original	Compacted		
	Top Lift				Bottom Lift			
Coarse aggregate No. 10 to 3 in.	56	56	54	54	57	55	54	53
Coarse sand No. 40 to No. 10	21	22	21	21	20	20	21	21
Fine sand No. 200 to No. 40	16.1	14.4	16.8	17.0	16.9	17.5	17.9	18.4
Silt 0.005 mm to No. 200	3.7	5.0	5.1	5.1	4.0	4.5	4.5	4.6
Clay < 0.005 mm	3.2	2.6	3.1	2.9	2.1	3.0	2.6	3.0
Plasticity index	N.P.		N.P.	N P.	N.P.	N.P.	N.P.	N.P.

Particle Size and Description	Test Section 2—Crushed Dolomite (Niagara)					
	Sampled During Construction[a]					
	Original	Compacted	Original	Compacted	Original	Compacted
	Top Lift		Center Lift		Bottom Lift	
Coarse aggregate No. 10 to 3 in.	76	70	70	67	73	72
Coarse sand No. 40 to No. 10	10	13	12	14	11	12
Fine sand No. 200 to No. 40	4.5	5.4	6.4	6.3	6.1	6.2
Silt 0.005 mm to No. 200	7.1	7.9	8.7	9.1	7.1	6.8
Clay < 0.005 mm	2.4	3.7	2.9	3.6	2.8	3.0
Plasticity index	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.

Particle Size and Description	Test Section 3—Dolomite (Prairie du Chien)				Test Section 5—Igneous Gravel			
	Sampled During Construction[a]		2 Years[b]	5 Years[b]	Sampled During Construction[a]		2 Years[b]	5 Years[b]
	Original	Compacted			Original	Compacted		
	Single Lift				Single Lift			
Coarse aggregate No. 10 to 3 in.	76	69	67	70	55	49	52	53
Coarse sand No. 40 to No. 10	8	9	11	10	25	27	26	24
Fine sand No. 200 to No. 40	7.9	10.3	10.8	9.3	14.3	15.7	15.2	14.8
Silt 0.005 mm to No. 200	6.1	8.6	8.9	8.5	3.7	5.2	5.3	5.4
Clay <0.005 mm	2.0	3.1	2.3	2.2	2.0	3.1	2.5	2.8
Plasticity index	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.

Particle Size and Description	Test Section 4—Igneous Gravel							
	Sampled During Construction[a]		2 Years[b]	5 Years[b]	Sampled During Construction[a]		2 Years[b]	5 Years[b]
	Original	Compacted			Original	Compacted		
	Top Lift				Bottom Lift			
Coarse aggregate No. 10 to 3 in.	62	63	62	58	57	58	58	58
Coarse sand No. 40 to No. 10	21	20	19	22	22	22	21	21
Fine sand No. 200 to No. 40	10.8	10.3	11.5	12.0	14.2	13.2	13.3	13.5
Silt 0.005 mm to No. 200	4.2	4.7	4.7	5.4	4.4	4.0	5.3	4.7
Clay <0.005 mm	2.0	2.0	2.8	2.6	2.4	2.8	2.4	2.8
Plasticity index	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.	N.P.

[a]Average of 30 samples, 25 lb each: Original, sampled before compaction; compacted, sampled after compaction.
[b]Average of 12 samples, 75 lb each: Sampled after 2 and 5 years of service.

TABLE 6

MEASURES OF DEGRADATION, TEST SECTIONS 1, 3, 4, AND 5

Method of Analysis	Test Section	Sampled During Construction[a]			Two Years[b]		Five Years[b]	
		Original	Compacted					
			Result	Percent Change[c]	Result	Percent Change[c]	Result	Percent Change[c]
Test Sections 3 and 5—Single Lift								
P-200 (percent)	3	8.1	11.7	+44	11.2	+38	10.7	+32
	5	5.7	8.3	+46	7.8	+37	8.2	+44
Surface area (sq cm/gm)	3	64	90	+40	87	+36	82	+28
	5	56	74	+32	71	+27	72	+28
Test Sections 1 and 4—Top Lift								
P-200 (percent)	1	6.9	7.6	+10	8.2	+19	8.0	+16
	4	6.2	6.7	+ 8	7.5	+21	8.0	+29
Surface area (sq cm/gm)	1	68	71	+ 4	77	+13	76	+12
	4	56	58	+ 4	64	+14	67	+20
Test Sections 1 and 4—Bottom Lift								
P-200 (percent)	1	6.1	7.5	+23	7.1	+16	7.6	+25
	4	6.8	6.8	0	7.7	+13	7.5	+10
Surface area (sq cm/gm)	1	66	74	+12	71	+ 8	73	+11
	4	63	65	+ 3	68	+ 8	66	+ 5

[a]Based on 30 samples, 25 lb each: Original, sampled before compaction; compacted, sampled after compaction.
[b]Based on 12 samples, 75 lb each: Sampled after 2 and 5 years service.
[c]Percent change from original value: + = percent increase; − = percent decrease.

Average gradations for samples obtained before compaction, after compaction, and after two and five years of service, are referred to as original, compacted, two-year, and five-year gradations respectively.

Particle sizes are regrouped into soil texture classifications in Table 5. The values listed in Table 5 for the coarse aggregate, coarse sand, and fine sand are based on the average values listed in Table 4, and the values for the silt and clay are based on hydrometer tests conducted on three representative samples for each condition.

Values of P-200 and surface area are given in Table 6 for test sections 1, 3, 4, and 5. These results are grouped according to similarity of construction, that is, number of layers required. Because test section 2 was sampled only during construction, the P-200 and surface area data are given separately in Table 7.

TABLE 7

MEASURES OF DEGRADATION, TEST SECTION 2

Method of Analysis	Sampled During Construction[a]		
	Original	Compacted	
		Result	Percent Change[b]
Top Lift			
P-200 (percent)	9.5	11.6	+22
surface area (sq cm/gm)	67	82	+22
Center Lift			
P-200 (percent)	11.6	12.7	+ 9
surface area (sq cm/gm)	83	91	+10
Bottom Lift			
P-200 (percent)	9.9	9.8	− 1
surface area (sq cm/gm)	72	73	+ 1

[a]No sampling subsequent to construction. Based on 30 samples, 25 lb each: Original, sampled before compaction; compacted, sampled after compaction.
[b]Percent change from original value: + = percent increase; − = decrease.

DISCUSSION OF RESULTS

Changes in the Distribution of Particle Sizes

Changes in the distribution of particle sizes with time (Table 5) generally occurred as a reduction of the coarse aggregate size and an increase in the fine sand and silt sizes for each test section. The magnitude of these changes varied between test sections and between lifts of the same test section. For example, compaction of the bottom lift of test section 2 produced very little change in the distribution of the particle sizes; compaction of the middle lift resulted in a decrease of the coarse aggregate size and small increses in the coarse sand silt and clay sizes; and compaction of the top lift resulted in a greater decrease in the coarse aggregate size and more pronounced increases in the coarse sand, fine sand, silt, and clay sizes.

The pattern of change in the particle size distributions for the top lifts of test sections 1 and 4 was quite similar. There was relatively no change in the amounts of coarse aggregate or coarse sand either after compaction or after two and five years of service. The amounts of fine sand increased slightly after two and five years, and the amounts of silt particles increased after compaction, but remained about the same after that. There was no important change in the clay content at any time.

There was some dissimilarity of changes in the particle size distributions of the bottom lifts of test sections 1 and 4. These dissimilarities occurred for the coarse aggregate and fine sand sizes. The amounts of coarse aggregate continually decreased with each subsequent sampling for test section 1, but the amount remained almost unchanged for test section 4. There was a slight increase in the amount of fine sand with each subsequent sampling of test section 1, compared to a 1 percent decrease after compaction for test section 4, with no subsequent change after two and five years. The amounts for the coarse sand, silt, and clay sizes varied somewhat for the two test sections, but there were no important changes in the amounts between the subsequent samplings.

The trends of change in the distribution of particle sizes for the single lifts of test sections 3 and 5 were similar to each other. There was an overall decrease in the amounts of coarse aggregate during the five years, relatively no change in the coarse sand sizes, overall gains in the amounts of fine sand and silt, and a small overall increase in the amount of clay.

The results show that the amounts of clay-sized particles were greatest after compaction for test sections 3 and 5, with lesser amounts for the subsequent two and five years. Although variability in the results was expected, this pattern of change was considered unusual. However, review of the field notes showed that there were delays of several days between the completion of compaction of these two sections and the after-compaction sampling, whereas there were no delays of sampling for the top lifts of test sections 1 and 4. Although the traffic during the short period of delay was minimal for test sections 3 and 5, it is possible that enough fine dust was generated and collected on the surface of the base course to produce results 0.5 to 0.9 percent higher for the compacted samples. The subsequent two- and five-year samples would not have contained this surface dust because the top ½ in. of surface of the base was removed before sampling so that the sample would not contain asphaltic material that would hinder sieving operations. As will be shown later, the small differences involved had no measurable influence on the interpretation of whether actual degradation occurred.

Perhaps the most important observations to be made from the changes in particle size distributions is that increases in clay-sized particles never exceeded 1.5 percent, and that the fines were never found to be plastic.

Magnitudes of Degradation

As mentioned previously, the pattern of degradation can be established from gradation curves and changes in the amounts of the various-sized particles, but the magnitude of degradation is best established from changes in single values, such as P-200 and surface area values. In addition, the percentage of change in these values allows a convenient means to assess the magnitude of change in these degradation values caused by compaction and subsequent service. It is evident from a comparison of percentage of change values in Table 6 that more degradation occurred in the single lifts of test sections 3 and 5 than in the top lifts of test sections 1 and 4 after five years of service. Also, it appears that compaction of the bottom lifts of test sections 1 and 4 produced some degradation, but that subsequent service did not produce much additional degradation.

Although the percentages of change provide a good comparison of the overall magnitude of degradation, they do not provide a convenient assessment of the pattern or trend of change in these values from one sampling period to the next. A better comparison was achieved by plotting each of the surface area and P-200 values in bar graph form.

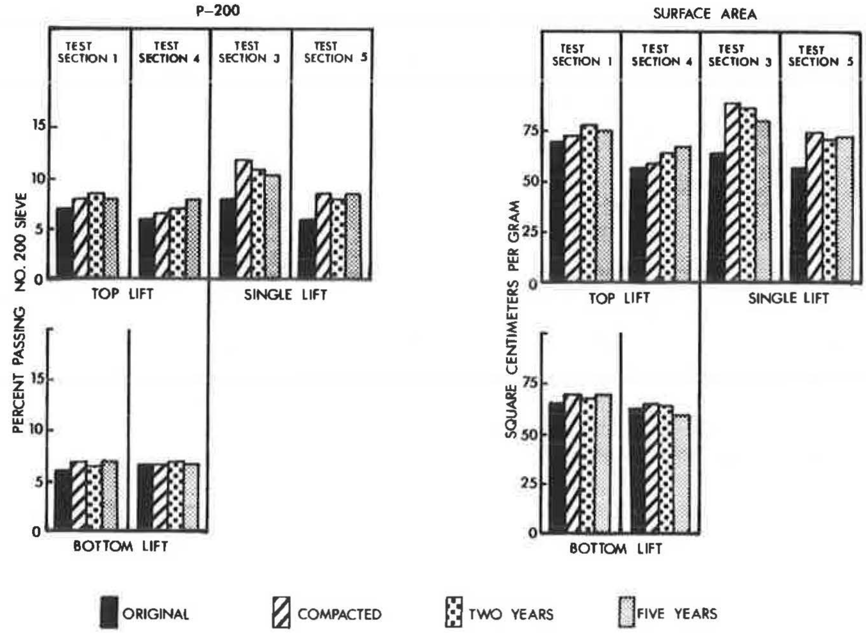

Figure 2. Measures of degradation.

These graphs (Fig. 2) show in general that (a) a small but gradual increase in the values took place during the five-year sampling period for the top lifts of test sections 1 and 4; (b) these values increased slightly because of compaction of the bottom lifts of test sections 1 and 4, but no further changes occurred; and (c) preponderant changes occurred because of compaction of the single lifts of test sections 3 and 5, but subsequent service produced no important additional changes. The possible reason for the slightly higher values of the compacted samples of test sections 3 and 5 was explained previously in the discussion of particle size distributions.

Although the degradation measures and the percentages of change in these measures indicated that degradation did occur, many of the changes were small and there were decreases as well as increases in values from one sampling period to the next. It was therefore believed that a statistical analysis that would compare one value to another and decide if these values differed statistically would provide a good guideline for assessment of degradation.

Statistical Analyses

Basically, the statistical analyses compared one average value to another and, depending on the variation among the individual results used to arrive at the average values, determined if the average values were statistically different. Only the percent-passing values for the No. 40 and No. 200 sieves were used, because the finer-sized material had the most influence on the magnitude of the P-200 and surface area values used for evaluation of degradation.

The basis of the comparison (5) was to determine whether the difference of the averages ($\overline{X}_A - \overline{X}_B$) was greater or less than a statistic (μ). The value of μ was dependent on the variability of the observations. If the difference was greater than the statistic ($\overline{X}_A - \overline{X}_B > \mu$), it was concluded that the averages being compared were different; otherwise, there was no reason to conclude they differed. As stated in Table 8 a "Yes" result indicates there is reason to conclude that the two average values being compared are different and, conversely, a "No" results means there is no reason to conclude that there is a difference.

TABLE 8

STATISTICAL COMPARISONS OF AVERAGE GRADATIONS—RESULTS OF THE ANALYSES

Comparison Between Gradations	Sieve Size, Percent Passing	Test Section 1			Test Section 4			Test Section 3			Test Section 5		
		$\bar{X}_A - \bar{X}_B$	μ	Yes or No[a]	$\bar{X}_A - \bar{X}_B$	μ	Yes or No[a]	$\bar{X}_A - \bar{X}_B$	μ	Yes or No[a]	$\bar{X}_A - \bar{X}_B$	μ	Yes or No[a]
					Top Lifts								
Original vs compacted	No. 40	0.6	1.6	No	0.7	1.2	No	6.1	1.4	Yes	4.5	1.6	Yes
	No. 200	0.5	0.8	No	0.5	0.5	No	3.6	0.7	Yes	2.6	0.8	Yes
					Bottom Lifts								
Original vs compacted	No. 40	2.1	1.2	Yes	0.2	1.3	No						
	No. 200	1.4	0.6	Yes	0.0	0.6	No						
					Top Lifts								
Compacted vs 2 years	No. 40	2.6	2.5	Yes	1.7	1.5	Yes	0.4	4.2	No	1.5	2.3	No
	No. 200	0.6	1.3	No	0.8	0.8	No	0.5	1.0	No	0.5	1.1	No
					Bottom Lifts								
Compacted vs 2 years	No. 40	0.3	1.5	No	0.7	1.3	No						
	No. 200	0.4	0.6	No	0.9	0.9	No						
					Top Lifts								
2 years vs 5 years	No. 40	0.0	3.1	No	1.0	2.3	No	2.0	2.6	No	0.0	2.6	No
	No. 200	0.0	1.5	No	0.5	0.9	No	0.5	1.3	No	0.4	1.3	No
					Bottom Lifts								
2 years vs 5 years	No. 40	1.0	2.9	No	0.2	1.2	No						
	No. 200	0.3	0.8	No	0.2	0.8	No						
					Top Lifts								
Compacted vs 5 years	No. 40	2.6	2.3	Yes	2.7	2.1	Yes	1.6	2.2	No	1.3	1.9	No
	No. 200	0.4	1.2	No	1.3	0.7	Yes	1.0	1.1	No	0.1	1.1	No
					Bottom Lifts								
Compacted vs 5 years	No. 40	1.3	2.8	No	0.7	1.3	No						
	No. 200	0.3	0.6	No	0.7	0.6	Yes						

[a]Yes—conclude there is a difference; No—conclude there is no difference.

The results of the analyses show that the compacted average P-40 and P-200 values differed from the original average values for test sections 3 and 5, but that subsequent average values did not differ from each other or the compacted values. These results were therefore interpreted as indicating that degradation did occur in the single lifts of test sections 3 and 5 between placement and compaction, but that no subsequent significant degradation occurred during the five years of service.

In contrast, the results for the top lifts of test sections 1 and 4 show only one instance where both the P-40 and P-200 values were statistically different. This was the comparison between the compacted and five-year values for the top lift of test section 4. These results were interpreted as indicating that no significant degradation occurred in the top lift of test section 1, but that gradual degradation took place in the top lift of test section 4 so that after five years of service a small but measurable amount of degradation had occurred.

Results for the bottom lifts of test sections 1 and 4 show that only the average original and compacted values of test section 1 differed. Consequently, compaction of the bottom lift of test section 1 produced some degradation, but subsequent service did not. No degradation occurred at any time within the bottom lift of test section 4.

It would appear from the evaluation of magnitude of degradation and the statistical comparisons that degradation of the aggregates in the five test sections occurred primarily because of breakage of the aggregate during compaction operations. The effect of wear during subsequent service was minimal. These data also show that the original gradation did not have a definite influence on the amount of degradation, as reported by other investigators (1). This is illustrated by a comparison between the igneous aggregates of test sections 4 and 5. The original gradations for the top and bottom lifts of test section 4 were almost the same as for the single lift of test section 5, but almost no degradation occurred in either lift of test section 4 caused by compaction, compared to a 32 percent increase of degradation in test section 5 (Table 6).

Effect of Type of Aggregate

Evaluation of the limited results obtained from this study did not indicate definite relationships between the types of aggregate and degradation. For example, the performances of similar igneous gravels in test sections 4 and 5 were quite opposite; compaction operations produced measurable degradation of the igneous gravel in test section 5, but subsequent service did not produce additional degradation. In contrast, compaction operations had no measurable influence on either lift of igneous gravel in test section 4, but the five years of service eventually produced some increase of fines in the top lift. Similar opposition of performances occurred for the dolomitic materials in test sections 1 and 3.

The limited results indicate that the thickness of the base course might have an influence on resultant degradation. The performance of the single lifts of test sections 3 and 5, compared respectively to the double lifts of the similar types of aggregate in test sections 1 and 4, shows that a greater degree of degradation occurred in the single lifts, and that compaction during construction produced this degradation.

No significant relationships could be identified between physical properties, such as soundness and wear, and the amount or occurrence of degradation. The field sampling crew did not report any evidence of yielding base courses during the five-year sampling operations. Consequently, the service performance of the selected base courses was considered satisfactory for the duration of the study.

OBSERVATIONS AND CONCLUSIONS

The following observations and conclusions appear justified from an evaluation of the test data obtained for the base aggregates incorporated in this study:

1. The relatively low percentages of clay-size particles in the as-produced base aggregates (less than 3.5 percent) were only slightly increased (less than 1.5 percent) by manipulation, compaction, or five years of service exposure.
2. Neither manipulation nor service exposure generated fines that changed the nonplastic nature of the as-produced aggregates.
3. Changes in the particle size distributions of the aggregates generally resulted in a reduction in the coarse aggregate sizes and an increase in the fine sand and silt sizes.
4. The greatest amount of degradation occurred during manipulation and compaction, and was greater for materials placed in single lifts than in multiple lifts. Five years of service exposure did not increase degradation significantly.
5. The primary cause of degradation appears to be breakage during manipulation and compaction rather than attrition during service exposure.
6. No relationship was indicated between magnitude of degradation and aggregate type or physical properties.

ACKNOWLEDGMENTS

This study was initiated by a discussion between J. R. Schultz, Chief Materials Engineer of the Division of Highways, Wisconsin Department of Transportation, and representatives of the Bureau of Public Roads concerning the desirability of such an investigation. Mr. Schultz's continuing interest, guidance, and review are indeed appreciated.

The assistance of District Materials Engineers and the respective County Highway Departments is gratefully acknowledged.

Special thanks are also expressed to the personnel of the Materials Research Unit who were so conscientious in conducting the field sampling, the arduous sieving operations, and the tabulation of results.

REFERENCES

1. Moavenzadeh, F., and Goetz, W. H. Aggregate Degradation in Bituminous Mixtures. Highway Research Record 24, pp. 106-137, 1963.

2. Aughenbaugh, N. B., Johnson, R. B., and Yoder, E. J. Information on Aggregate Degradation—A Literature Review. Civil Engineering School Rept., Purdue Univ., 1961.
3. Minor, C. E. Degradation of Mineral Aggregates. ASTM, Spec. Tech. Publ. 277, 1959.
4. Degradation of Aggregates Used in Highway Base Construction. HRB Bull. 344, 1962.
5. Experimental Statistics. National Bureau of Standards Handbook 91, 1963.