AHTD TRC TRANSPORTATION RESEARCH COMMITTEE

TRC0306

Alternative Test Methods for Air Voids and Field Density

Stacy G. Williams

Final Report

2006

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle

TRC-0306
Alternative Test Methods for Air Voids and Field Density
Final Report

5. Report Date
April 2006

6. Performing Organization Code

7. Author(s)
Stacy G. Williams

8. Performing Organization Report No.

9. Performing Organization Name and Address
University of Arkansas, Department of Civil Engineering
700 Research Center Blvd.
Fayetteville, AR 72701

10. Work Unit No.
(TRAIS)

11. Contract or Grant No.
TRC-0306

12. Sponsoring Agency Name and Address
Arkansas State Highway and Transportation Department
P.O. Box 2261
Little Rock, AR 72203-2261

13. Type of Report and Period covered
Final Report
1 Jan 03 thru 30 Jun 04

14. Sponsoring Agency Code

15. Supplementary Notes
Conducted in cooperation with U.S. Department of Transportation,
Federal Highway Administration

16. Abstract

The measurement of specific gravity for hot-mix asphalt (HMA) is critical in almost every topic pertaining to asphalt mixtures. Thus, being able to accurately and precisely measure this property is of paramount importance. During design and construction, the bulk specific gravity (Gmb) and maximum theoretical specific gravity (Gmm) of HMA mixes are used to calculate most of the properties that indicate pavement quality.

In this study, alternative methods for Gmb and Gmm were evaluated and compared to traditional measures of these properties to assess precision and relative accuracy. Three methods for Gmm were considered (CoreLok, Kuss, and AASHTO T-209) and evaluated with regard to the effects of nominal maximum aggregate size (4 sizes were tested – 9.5mm, 12.5mm, 25.0mm, and 37.5mm) for a selection of aggregate types typically found in Arkansas (4 sources were tested – sandstone, syenite, gravel, and limestone). Five methods for Gmb were considered (CoreLok, CoreReader, Height-Diameter, Kuss, and SSD) and evaluated for the same sources and sizes, as well as level of compactive effort (3 levels were tested – high, medium, and low).

The results indicate that, in terms of precision, the traditional methods exhibit the lowest levels of variability. Accuracy can only be assessed relatively, but it was determined that NMAS and compactive effort significantly affect test results. For Gmb measurements, the Height-Diameter method is most sensitive to changes in NMAS, while the CoreLok and Kuss methods are least sensitive to these changes. In most cases, CoreLok and SSD values are similar for small NMAS mixes, but not for large NMAS mixes.

Strong mathematical correlations were developed in order to relate traditional and alternative Gmb test methods based on NMAS. The strongest and most practical relationships were developed between the SSD and CoreLok methods. These models can be used to assess the impacts of incorporating new test methods for both design and construction procedures. They can also be used to normalize data from different test methods, should alternative methods be incorporated into the current specification.

While alternative methods do possess significant advantages, the results of this study do not support the elimination of traditional methods.

17. Key Words
Asphalt, Asphalt Mixtures, Air Voids, Density
Bulk Specific Gravity

18. Distribution Statement
No Restrictions

19. Security Classif. (Of this report)	20. Security Classif. (Of this page)	21. No. of Pages	22. Price
(none)	**(none)**	**93**	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ALTERNATIVE TEST METHODS FOR AIR VOIDS AND FIELD DENSITY

TRC-0306

by

Stacy G. Williams, Ph.D., P.E.
Research Assistant Professor
Department of Civil Engineering
University of Arkansas
700 Research Center Blvd.
Fayetteville, AR 72701
479-575-2220
479-575-7639 (FAX)
sgwill@uark.edu

April 2006

ABSTRACT

The measurement of specific gravity for hot-mix asphalt (HMA) is critical in almost every topic pertaining to asphalt mixtures. Thus, being able to accurately and precisely measure this property is of paramount importance. During design and construction, the bulk specific gravity (Gmb) and maximum theoretical specific gravity (Gmm) of HMA mixes are used to calculate most of the properties that indicate pavement quality.

In this study, alternative methods for Gmb and Gmm were evaluated and compared to traditional measures of these properties to assess precision and relative accuracy. Three methods for Gmm were considered (CoreLok, Kuss, and AASHTO T-209) and evaluated with regard to the effects of nominal maximum aggregate size (4 sizes were tested – 9.5mm, 12.5mm, 25.0mm, and 37.5mm) for a selection of aggregate types typically found in Arkansas (4 sources were tested – sandstone, syenite, gravel, and limestone). Five methods for Gmb were considered (CoreLok, CoreReader, Height-Diameter, Kuss, and SSD) and evaluated for the same sources and sizes, as well as level of compactive effort (3 levels were tested – high, medium, and low).

The results indicate that, in terms of precision, the traditional methods exhibit the lowest levels of variability. Accuracy can only be assessed relatively, but it was determined that NMAS and compactive effort significantly affect test results. For Gmb measurements, the Height-Diameter method is most sensitive to changes in NMAS, while the CoreLok and Kuss methods are least sensitive to these changes. In most cases, CoreLok and SSD values are similar for small NMAS mixes, but not for large NMAS mixes.

Strong mathematical correlations were developed in order to relate traditional and alternative Gmb test methods based on NMAS. The strongest and most practical relationships were developed between the SSD and CoreLok methods. These models can be used to assess the impacts of incorporating new test methods for both design and construction procedures. They can also be used to normalize data from different test methods, should alternative methods be incorporated into the current specification.

While alternative methods do possess significant advantages, the results of this study do not support the elimination of traditional methods.

ACKNOWLEDGMENTS

The author wishes to sincerely acknowledge the support of the Arkansas State Highway and Transportation Department (AHTD) for funding this research adventure and for coordinating the acquisition of samples and materials. The talents of Artem Popov, John T. Krentz, Nhung Le, Rajesh Janarthanan, and Frances Griffith are also greatly appreciated. The personnel of AHTD RE 43, APAC-Arkansas (McClinton-Anchor and Arkhola Divisions), Granite Mountian Quarries, and Jet Asphalt, Inc. are also recognized for their efforts and support.

The contents of this paper reflect the views of the author who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration or the Arkansas State Highway and Transportation Department. This paper does not constitute a standard, specification, or regulation; it should not be considered an endorsement of any commercial product or service.

INTRODUCTION

The measurement of specific gravity for hot-mix asphalt (HMA) is critical in almost every topic pertaining to asphalt mixtures. Thus, being able to accurately and precisely measure this property is of paramount importance. In design, the bulk specific gravity, or density, of a compacted mixture (Gmb) is used for the determination of volumetric properties such as air voids, voids in the mineral aggregate (VMA), voids filled with asphalt (VFA), and percent density after compaction. These properties are associated with specification requirements for asphalt mixture design as outlined in AASHTO M-323, as well as the *Standard Specification for Highway Construction* for the Arkansas State Highway and Transportation Department (AHTD). *(1, 2)* During construction, the property of bulk specific gravity is used in most measures of quality control and quality assurance (QC/QA). Problems with asphalt density have been linked to pavement distresses such as rutting, stripping, bleeding, cracking, age hardening, and excessive permeability. *(3, 4)* In Arkansas, the bulk specific gravity of compacted asphalt materials is a major component in calculating three of the four pay items for the contractor. Thus, errors in this measurement have very real consequences in terms of both contractor pay and pavement quality.

The measure of maximum theoretical specific gravity of the mix (Gmm) is also a critical property for HMA materials. This value represents the density of the mixture as if it were compressed into a solid mass of aggregate and asphalt cement. The ratio of bulk specific gravity to the maximum theoretical specific gravity represents the percent density, and is also used to calculate percent air voids. Percent air voids and percent density in the field are both pay items in Arkansas. Thus, errors in this measurement can also have potentially serious consequences in terms of contractor compensation and pavement quality. Additionally, because both Gmb and Gmm are used in calculating these quantities, the potential overall error is compounded.

Accurate and precise methods for measuring both Gmb and Gmm are needed in order to improve the quality and consistency of HMA mixture design and construction. Several methods for quantifying each of these quantities are investigated in this research project.

BACKGROUND

Bulk specific gravity is a measure of density, which requires a measure of both weight and volume. Accurate weights are easily obtained, but accurate volumes are a bit more elusive, especially for samples that possess an irregular surface texture. Several methods will be examined in this study. The following is a discussion of each of the methods included.

SSD Method

Traditionally, the bulk specific gravity of compacted asphalt mixes (Gmb) has been measured using the water displacement, or SSD method. This procedure, outlined in AASHTO T-166 and ASTM D2726 *(1)*, involves first weighing the dry sample in air, then submerging it in water at 25 C for a period of three to five minutes, and recording the submerged weight. Then the sample is removed from the water bath, brought to the saturated surface-dry (SSD) condition, and the SSD weight is recorded. This procedure is based on Archimedes' principle that the volume of an object placed in water is equal to the volume of water displaced by that object. The equation used to calculate Gmb is given in Equation 1.

$$Gmb = \frac{Dry\ Mass}{(SSD\ Mass - Submerged\ Mass)} \qquad \textit{Equation 1}$$

The SSD method has several advantages in that it is relatively inexpensive and simple to perform. Also, it is the traditional standard upon which most asphalt mixture design procedures and QC/QA specifications are based. The greatest disadvantage of this method is that for coarse-graded mixes, an accurate SSD weight can be difficult to achieve. Also, such mixes are prone to larger, more interconnected void pathways. This interconnectivity can allow water to enter the interior portion of the sample, thereby underestimating the sample volume, especially if this water drains from the sample before the SSD weight measurement can be secured. As the volume is underestimated, the Gmb is overestimated, and in turn, calculated air voids for the sample are then underestimated. This concept has become more evident as recent mix design specifications have shifted toward more coarse-graded mixes. *(2, 5)* The popularity of such mixes has prompted the investigation of other test methods to serve as alternatives to the SSD method.

Height-Diameter Method

The Height-Diameter method is based on a dimensional analysis and is discussed in AASHTO T-269 *(1)*. The height and diameter of a compacted asphalt sample are measured, each at four evenly spaced locations around the cylinder. The average height and average diameter are then used to calculate the specimen volume. By obtaining the dry weight of the sample, a direct ratio of weight to volume can then be calculated in order to determine Gmb. The greatest advantage of this method is that it is extremely simple, quick, and inexpensive. After testing, the sample is unaltered and can be used for other purposes. The greatest disadvantage is that this method assumes the asphalt specimen to be a perfect cylinder, which maximizes the surface voids included in the bulk volume. This phenomenon is more exaggerated for coarse-graded mixtures possessing a large number of surface irregularities. As a result, the Height-Diameter method typically underestimates Gmb, thereby overestimating air voids.

CoreLok Method

The CoreLok, shown in Figure 1, is a relatively new test method that has gained popularity in recent years. In this method, a compacted asphalt sample is vacuum-sealed in a specially designed plastic bag using an automatic vacuum chamber. The plastic bag conforms to the surface of the sample and prevents water infiltration. After the weights of the sample and bag have been determined in air, the sample is vacuum-sealed, and a submerged weight is determined. The resulting Gmb is then corrected for bag density and volume. The concept is that the bag is strong enough to seal the sample without puncturing, while pliable enough to conform to the surface irregularities of the sample. In this way, the measured volume of a coarse-graded mixture with many interconnected voids is believed to be more accurate because it is neither overestimated by being assumed to be a perfect cylinder nor underestimated by allowing water to penetrate the interior portions of the sample. The complete method for this test is outlined in ASTM D6752, and is summarized as follows:

- Determine the dry mass of the unsealed HMA sample.
- Place the plastic bag in the sample chamber, then insert the sample into the bag.
- Close the vacuum chamber. The device will automatically begin the vacuum-sealing process, evacuating the chamber to 760 mm Hg.
- When the vacuum-sealing process is complete, the chamber door will open, revealing the sealed sample.
- Record the weight of the sealed sample in air, then record its weight submerged in water.

- Calculate the bulk specific gravity of the compacted HMA sample, correcting the results for the bag density.

Figure 1. The CoreLok Device

Advantages of the CoreLok method are that it is relatively simple to perform, it can test highly absorptive samples, and upon completion of the test, the sample is dry and can be used for other purposes. A complete test takes approximately five minutes. The primary disadvantage is that the bags are single-use items, creating a recurring expense to the user. Also, if a bag is punctured during the test, water is able to enter the sample, and the test results must be discarded. The sample cannot be re-tested until the water is removed from the specimen, and in some cases, the sample may never regain its completely dry status.

CoreReader Method

The CoreReader, shown in Figure 2, was developed by Troxler Electronic Laboratories, Inc. as a way to directly measure the density of HMA samples. The CoreReader measures bulk specific gravity directly using a low level gamma ray source, much like that of the nuclear gauges which are routinely used for field density determinations. *(6)* This eliminates the need for an approximation of sample volume. In this method, the asphalt sample is placed on the sample tray in the CoreReader and a sample height is entered. After approximately seven minutes, a Gmb measurement is provided.

Figure 2. The CoreReader Device

An advantage of this method is that it involves extremely minimal operator effort and interaction, and is completely nondestructive to the specimen. This device, when properly calibrated, is reported to provide repeatable and accurate measurements that are not operator dependent. The primary disadvantage of this method is the expense of the testing unit. *(6, 7)*

Kuss Method

The Kuss method for the measurement of Gmb, shown in Figure 3, was developed by Mr. Mark Kuss at the University of Arkansas. In this method, the dry weight of an asphalt sample is recorded, then the specimen is placed in a partially-filled water column, causing the water level to rise. The pressure change above the water column is measured by a patented pressure measurement system, and the bulk volume of the specimen is estimated in approximately fifteen seconds. The mass and volume are then used to calculate the Gmb using traditional mass-volume relationship. The most advantageous feature of this method is that the volume of the sample is determined very quickly. This not only allows the operator to generate results quickly, but also estimates sample volume before the water has a chance to absorb into the interior portions of the sample by way of interconnected void pathways. Thus, the error in underestimating volume, as documented for the SSD method, is reduced. Also, the sample does not have to be removed from the water or brought to the SSD condition in order to estimate volume. An entire test can be completed in less than two minutes. The most notable disadvantage to this method is that after testing, the sample must be dried if the specimen is to be used for other purposes.

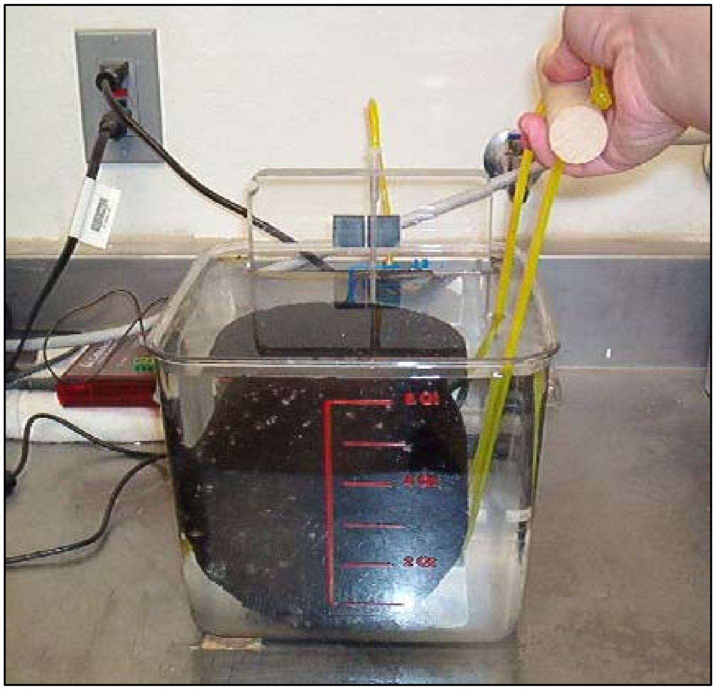

Figure 3. The Kuss Gmb Device

Paraffin Method

Several other methods have been used for the determination of Gmb measurements of HMA mixtures. Most of these methods involve coating or sealing the exterior surfaces of the specimen in order to estimate the volume. One such method is the paraffin coating method, which is described in AASHTO T-275. *(1)* In this method, a dry sample is weighed, and then dipped in melted paraffin. As the paraffin cools, it forms a coating that seals the surfaces of the sample. Next, the coated sample is weighed in air and in the submerged state. A calculation similar to that of the SSD method, along with a correction for the specific gravity of the paraffin, is used to determine the density of the HMA sample. According to AASHTO T-166, any sample that exhibits a water absorption value in excess of 2.0 percent should be tested according to this method. There are several problems with this method. First, the procedure is time consuming and rather cumbersome to perform, increasing the chance for operator variability; and it is difficult to use the resulting sample for further testing. Also, since T-166 must first be performed to determine its absorption capacity, additional time is required for the sample to dry before it can be tested by the paraffin method, if it does in fact reach the dry state.

It is important to note that the SSD method has been scrutinized for its ability (or lack thereof) to estimate volume and SSD measurements. These measurements must be made in order to calculate absorption. So, one must also question the accuracy of the calculated absorption capacity. Thus, the determination of which samples should be tested according to AASHTO T-275 must also be considered suspect.

Parafilm Method

Parafilm is another method that has been used. This process involves wrapping the compacted HMA sample with parafilm, then testing in a manner similar to that discussed previously in the paraffin method. The method is fully described in ASTM D1188. The parafilm method combines the principles of the paraffin method with the ability to remove the material for a clean sample after testing, but there are disadvantages. The most notable problems associated with this method are difficulties in properly wrapping the samples and producing consistent results between operators. Additionally, the parafilm tends to “bridge” the voids of samples with surface irregularities, which overestimates the sample volume – much like that of the Height-Diameter method. *(3)*

Rice Method

The traditional standard for the measurement of the theoretical maximum specific gravity, or Gmm, is AASHTO T-209, which was developed by James Rice in the 1950s, and is commonly referred to as the Rice method. *(8)* In this method, a sample of loose HMA mix is weighed in air, then placed under vacuum for a period of approximately fifteen (15) minutes in order to remove the air. Finally, the sample is weighed in the submerged state. The maximum theoretical specific gravity is then calculated according to Equation 2.

$$Gmm = \frac{Dry\ Mass}{\left(Dry\ Mass - Submerged\ Mass\right)} \qquad \textit{Equation 2}$$

The Rice value is an apparent specific gravity rather than a bulk specific gravity, since it involves the volume of just the solid material without considering any pore volume.

Although this method has been used for many years, it has been scrutinized due to the large variation in methods and equipment that can be used for the test. *(9)* Thus, new methods for the measurement of this property have been proposed.

CoreLok Method

Although the CoreLok device was originally developed for the measurement of Gmb values, a procedure has also been developed for the measurement of Gmm. The method has undergone a series of revisions and improvements. The current procedure is summarized as follows.

A loose mix is weighed and placed inside a plastic bag within the vacuum chamber. The bag that is used for this method has a special "channeled" texture on one side. The weight of the sample in the bag is recorded after the sealing process. Next, the sealed sample is submerged and the bag is cut open while the sample is submerged. The difference in pressure allows water to infiltrate the sample and completely surround the particles in the sample without trapping air bubbles. The submerged weight of the sample and bag is recorded, and Gmm is calculated according to Equation 3.

$$Gmm = \frac{Dry\ Mass}{\left(Dry\ Mass - Submerged\ Mass\right)} \qquad \textit{Equation 3}$$

Kuss Method

The Kuss method has also been adapted for use in the determination of Gmm. This method, shown in Figure 4, was also developed by Mark Kuss. The procedure is similar to the Kuss Gmb method, but uses air rather than water since Gmm values are actually measures of apparent specific gravity. In this method, a sample of loose mix is placed in a canister of known volume and sealed. Then a measured quantity of air is introduced into the canister. The pressure differential is measured using the patented pressure measurement system, which translates to an apparent volume estimation. This value, along with the sample weight, is used to calculate the Gmm for the mixture based on the weight-volume relationship.

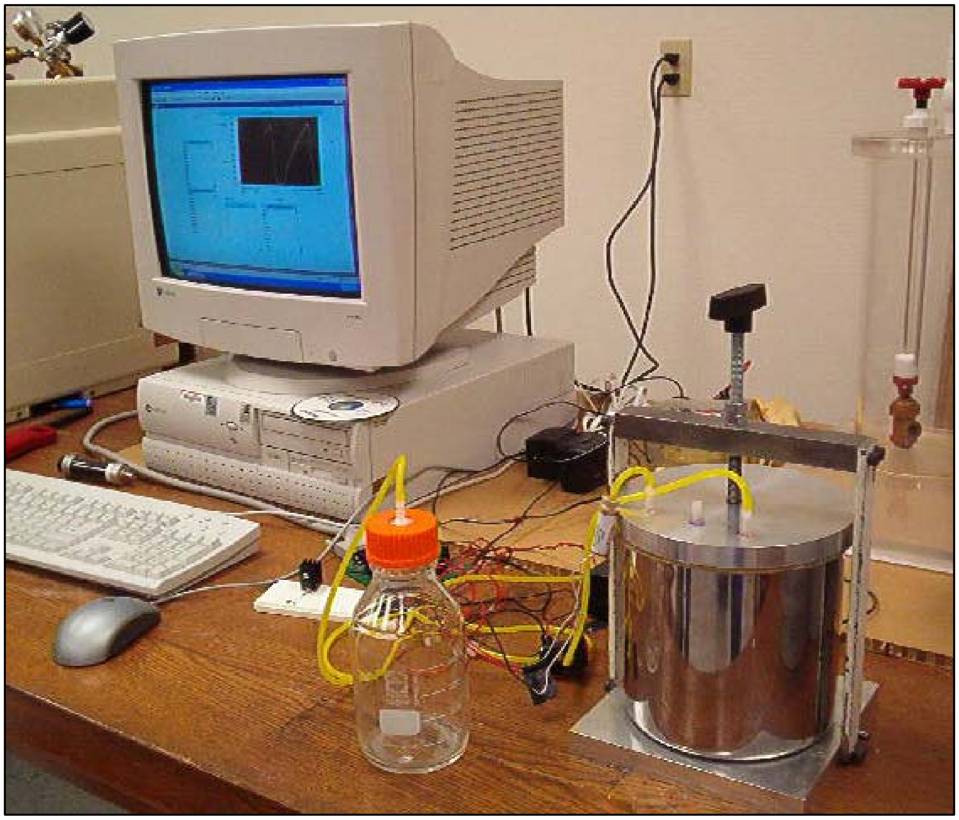

Figure 4. The Kuss Gmm Device

Air Voids

The air void content of a mixture is expressed as a percentage by volume. Equation 4 shows the relationship.

$$Air\ Voids = \left(1 - \frac{Gmb}{Gmm}\right) * 100 \qquad \textit{Equation 4}$$

Another form of this quantity is percent density, which is given by Equation 5. Most specifications are based on the values of air voids and percent density, which demonstrates the importance of Gmb and Gmm.

$$Percent\ Density = \left(\frac{Gmb}{Gmm}\right) * 100 \qquad \textit{Equation 5}$$

LITERATURE REVIEW

As the Superpave mixture design methods have been implemented and coarse-graded mixtures have become more popular, the topic of accurately determining Gmb values has received a considerable amount of attention. Several studies have been conducted in order to assess the accuracy and precision of various testing methods, with much attention devoted to the CoreLok vacuum sealing device. Several of these studies are discussed in this section.

Buchanan

In 2000, Buchanan reported on a study involving four mix types (fine-graded and coarse-graded Superpave, Open Graded Friction Course (OGFC), and Stone-Matrix Asphalt (SMA)), all having a nominal maximum aggregate size (NMAS) of 12.5mm. *(3)* Two aggregate types (limestone and granite) were used for each mix type, and samples were compacted to a range of three compactive efforts (low, medium, and high) representing a wide range of air void levels. Triplicate samples were prepared for each combination of testing parameters, and tested according to four methods (SSD, CoreLok, Height-Diameter, and Parafilm). One challenge that arose during the testing is the fact that a density gradient exists in gyratory-compacted samples. Because samples are compacted in gyratory molds with fixed walls, the surfaces of the sample are more irregular than that of a core cut from a field-compacted asphalt mat. Thus, the gyratory-compacted sample density near the edges is significantly less than that in the center portion of the sample. This phenomenon creates a discrepancy in measures of Gmb, especially when measured by the Height-Diameter method.

In order to assess the accuracy of the methods, a "true" measure of Gmb must be determined. But because there is no "true" measure of Gmb by which to compare the results, Buchanan decided that the Height-Diameter method could provide the truest measure if only the center portions of each sample were tested. So after the original testing, the samples were sawed into cubical shapes. By doing this, many of the surface irregularities were eliminated, and thus the Height-Diameter estimation for volume was assumed to be the most accurate. The samples were then tested in the "cut" condition, and compared to the results for the "uncut" condition. Cutting the samples resulted in lower testing variability between the various methods.

In 79 percent of the observations of fine- and coarse- graded Superpave mixes, there was no statistical difference between the SSD and CoreLok methods. When a statistical difference did exist, the CoreLok method produced lower densities. The height-diameter and parafilm methods were similar in over half the cases. It was also concluded that as the surface texture became rougher, the difference between the Height-Diameter and CoreLok methods increased. Although it was anticipated that the cut samples measured by the Height-Diameter method would provide the most accurate measure of Gmb, difficulties were encountered in sawing the samples. In several cases, the opposite faces were not truly parallel, and sample degradation was common at the end of the saw cuts. It was noted that a 1 percent error in the volume estimate for a 1200 gram sample can result in an error for calculated air voids of 1.0 percent. Thus, errors that arose due to the sawing process could have significantly affected the results of the height-diameter measurements, and thus may not have been the most accurate measures of Gmb.

Overall, the CoreLok method was chosen to be the most accurate, and was cited as the least affected by the testing matrix parameters.

Hall, Griffith, and Williams

A study at the University of Arkansas tested a total of 144 samples multiple times by multiple operators, using three test methods (SSD, CoreLok, and Height-Diameter). The samples were gyratory-compacted from field mix obtained from 24 separate stations on six jobs. All mixes were 12.5mm coarse-graded Superpave mixtures. *(10)*

First, the methods were compared. In 83 percent of the 24 cases, there was a statistically significant difference between methods. In 71 percent of the 24 cases, there was a statistically significant difference between the SSD and CoreLok methods, which is similar to the results reported by Buchanan. In general, the Height-Diameter method produced the lowest sample densities, and the SSD method produced the highest sample densities.

In terms of operator variability, the Height-Diameter method had the lowest average standard deviation for each of the 24 cases, but when comparing the replicate tests on individual samples, the CoreLok exhibited the smallest variability for over half the 144 samples. When comparing only SSD and CoreLok methods, the CoreLok variability was less than that of the SSD method in 82 percent of the 144 individual samples. The conclusion of this study included a recommendation for further study on the CoreLok as a viable alternative for the measurement of Gmb for HMA.

Crouch, et. al.

Another study conducted in Tennessee, focused on four test methods – SSD, CoreLok, Height-Diameter, and Parafilm. *(4)* In this study, the precision and accuracy of the chosen methods was measured by performing seven replicate tests by each method on a set of 50 widely varied HMA samples, and three replicate tests on a set of four aluminum samples.

The SSD method was shown to have the least variability, as described by coefficient of variation (COV). The COV values are summarized below in Table 1.

Method	SSD	CoreLok	Parafilm	Height-Diameter
COV (%)	0.08	0.20	0.26	0.34

Table 1. Comparison of variability of Gmb measurements by Tennessee Study *(4)*

In preliminary testing, the Height-Diameter method was less variable than both the CoreLok and Parafilm methods. However, in the full-scale testing matrix, it exhibited the greatest level of variability. It was noted that this was probably due to the fact that the Height-Diameter method was especially variable for field samples and for samples having any damage or surface irregularities. In general, it was concluded that all of the methods have relatively low variability, having less than 0.5% COV.

From a practical standpoint, it was concluded that SSD should not be used because it is not intended for use with open samples that have interconnecting voids or more than 2.0 percent water absorption. The Height-Diameter method was eliminated as a viable alternative due to its inability to handle samples with surface irregularities. Parafilm was also eliminated for this reason, as well as the impracticality of its time-consuming nature. By process of elimination, the CoreLok method was recommended for use as the most widely applicable to various sample types. However, its accuracy is unknown. In order to assess the accuracy of the CoreLok, the Gmb values obtained by the various methods were compared.

In general, the SSD method produced the highest densities, followed by the CoreLok, then Parafilm. The height-diameter method generated the lowest sample densities. In nine of ten sample groups, the difference in the CoreLok and SSD results were statistically significant. From this data, it was concluded that the SSD method formed an "upper" boundary for density measurements, while the height-diameter and parafilm methods formed a "lower" boundary. Thus, the CoreLok Gmb could be considered to be relatively accurate since they fell within the bounds of what could be considered to be a range of "true" values.

NCAT

Growing interest in the CoreLok method for Gmb measurement prompted a Round-Robin study using the device. *(11)* In this project, 18 laboratories performed testing to determine the repeatability and reproducibility for both the SSD and CoreLok methods. The repeatability, or within-laboratory variability, represents the ability of a single operator to generate consistent test results when repeatedly performing a given test method on a particular sample, using the same procedures and equipment. In this situation, all variables are held constant so that the measured variability can be assumed to be a result of the test method itself. The reproducibility, or between-laboratory variability, involves multiple laboratories performing the same test method on the same sample. In this case, additional variation is present that can be attributed to different operators and different laboratory equipment.

One aggregate type (quarried granite) was used to create three mixture types (SMA, coarse-graded and fine-graded Superpave). Each mixture was compacted, in triplicate, to each of three compactive efforts (low, medium, and high). Only one nominal maximum aggregate size (12.5mm) was tested.

In terms of variability, the SSD method had a lower standard deviation than CoreLok for both within-laboratory and between-laboratory measures. Based on F-tests, the variances of the SSD and CoreLok methods were similar for six of nine mixes. In the three mixes having variances that were statistically different, the SSD method exhibited less variability. A trend of decreasing variability was reported as the Gmb values increased for coarse mixes. This is reasonable because as densities increase, the amount of interconnected voids and surface irregularities decrease, thereby decreasing the variability of the test methods. Although the CoreLok variability was significantly higher than that of the SSD method, it was reported to be less affected by changes in mix type and air voids.

One of the primary results of this study was the development of a precision statement for the CoreLok method. “The single-operator standard deviation has been found to be 0.0124. Therefore, results of two properly conducted tests by the same operator on the same material should not differ by more than 0.035. The multi-laboratory standard deviation has been found to be 0.0135. Therefore, results from two properly conducted tests from two different laboratories on samples of the same material should not differ by more than 0.038.” AASHTO T 166 states that the results of two properly conducted SSD tests by the same operator on the same material should not differ by more than 0.02, which is considerably smaller than that for the CoreLok method. However, it was also noted that standard deviations for the SSD method in this research were too large for that value to be feasible.

The NCAT study also generated d2s values for both the SSD and CoreLok methods. *(11)* The results indicated that the within lab d2s value for the SSD method should actually be 0.052, which is much larger than that published in the AASHTO standard. *(1)* Results from several round robin studies were cited, listing within-lab d2s values ranging from 0.026 to 0.052.

In terms of a method comparison, the SSD and CoreLok measures of Gmb were almost identical for fine-graded mixes. However, significant differences were detected for the coarser mixes, with the CoreLok producing lower densities than the SSD method, especially at the low compactive effort. The differences were not constant, and varied with changes in gradation and compactive effort. These differences appeared to be sensitive to water absorption values, being more pronounced for samples having absorption values greater than 0.4%.

A final product of this study is the recommendation that the CoreLok method is a viable option for measurements of Gmb, and although its variability is slightly greater than that of the SSD method, it is believed to be more accurate for samples with higher absorption values.

AMRL

Another study was performed under NCHRP Project 9-26 to assess the variability of the SSD and CoreLok methods for measuring Gmb. *(12)* The results were reported in 2004 by members of the AASHTO Materials Reference Laboratory (AMRL).

This study involved 3 mix designs – a 19.0mm coarse-graded mix, a 12.5mm fine-graded mix, and a 9.5mm fine-graded mix. One binder (PG 64-22) and one aggregate source (limestone) was used. All samples were compacted to approximately three percent air voids, and had about 0.5 percent absorption. Nine samples were prepared for each mixture, and 20 laboratories were chosen to perform replicate tests on the specimens using the SSD method, with ten of those laboratories also performing CoreLok testing. The AMRL laboratory performed additional testing using the CoreReader.

The CoreLok was not as repeatable as the SSD method, which was proven even when aluminum cylinders were tested. The SSD was very precise, and exhibited a 0.07 percent COV for all three mixes. CoreLok variability by this measure was about three times that of the SSD method. Based on this low level of variability for the SSD method, it was concluded that there is no need to attempt to find ways to improve AASHTO T-166 for these types of samples.

For the SSD testing, variability was separated into three separate components which corresponded with variability of the specimen, variability of the laboratory, and variability of the method. This variability analysis revealed that very little of the variability in the SSD method was due to the test method itself. Most (approximately 90 percent) of the variability was attributed to the mixing and compacting process. The laboratory component added very little to the total estimate of variability.

Standard deviations for repeatability (within-laboratory variability) and reproducibility (between-laboratory variability) are shown in Table # for the three methods used in this research. Since the CoreReader was used in only one laboratory, there is no estimate for reproducibility.

	Repeatability	Reproducibility
SSD	0.002	0.003
CoreLok	0.006	0.007
CoreReader	0.004	NA

Table 2. Standard deviations associated with repeatability and reproducibility for various test methods.

Based on F-tests, the variability of all three methods was similar for the 9.5mm mixture. With respect to the 12.5mm mixture, only CoreLok and CoreReader showed similar variability. For the 19.0mm mixture, only the SSD and CoreLok methods were similar.

When comparing the Gmb measurements by the two methods, the SSD Gmb values were higher than the CoreLok Gmb values. This discrepancy appeared to increase as the amount of surface irregularities increased, which is expected as the NMAS of the mixture increases. Greater differences were exhibited by the CoreReader as compared to the SSD method.

Overall, this study concluded that the SSD method is a very good test for samples having 3.0 percent air voids and 0.5% absorption. This statement begs one to question the applicability of the SSD method for other types of samples.

Conclusion

Although the SSD method continues to exhibit the lowest levels of variability, there seems to be a common belief that the CoreLok method may be able to provide more accurate results, while being less affected by other mixture parameters.

OBJECTIVES

The overall objective of this project was to evaluate alternative testing methods for the determination of air voids and field density of HMA mixtures. Specific objectives follow.

Summarize current research efforts relative to new methods being used for the determination of Gmb and Gmm. Much research is currently ongoing around the country relating to new and innovative methods for the measurement of density and air voids of HMA mixtures. Because of the growing interest in this research topic, a summary of such projects will first be documented.

Investigate new methods for use in the measurement of air voids and field density of HMA mixes. As new methods are made available, their potential for standardized use should be evaluated.

Assess the variability of alternative testing methods relative to specific gravity measurements. One of the primary measures of effectiveness for a testing method is its level of variability. By minimizing the variability of a procedure, the measured variability can be attributed to the material rather than the procedure, making it a much more attractive method for testing in design and quality control/quality assurance situations.

Assess the relative accuracy of new and traditional testing methods. Since no absolutely true measure of specific gravity for HMA specimens exists, any comparisons of accuracy must be relative. Although no method can be deemed exact, all current specifications are based on the values obtained through the use of traditional methods. Therefore, any relative change in specific gravity measurements, as well as any impacts on existing specifications, must be considered before the recommendation of a new test method. If a new testing method is recommended for use, mathematical models correlating the resulting measurements of the new and traditional methods would be very beneficial.

Relate the effects of alternative testing methods to measured field densities according to the nuclear method. Laboratory specific gravity tests are used in determining job correction factors for nuclear density testing of compacted HMA pavements. Differences in measured Gmb and/or Gmm values, or the variability of such values, have a significant effect on this correction factor. Therefore, any impacts of alternative test methods on nuclear testing procedures must be investigated.

SCOPE

This research study investigated various methods used to measure the properties of both Gmb and Gmm. In order to assess the applicability of conclusions to a variety of aggregate types, a selection of aggregate sources was chosen to represent the typical range of materials found in the state of Arkansas. Four aggregate sources were selected including limestone (LS), sandstone (SS), gravel (GR), and syenite (SY). From each aggregate source, mixes were designed at four nominal maximum aggregate sizes. These sizes were 9.5mm, 12.5mm, 25.0mm, and 37.5mm, which comprise the four mixture sizes contained in the AHTD Construction Specification for the design of asphalt mixtures. *(2)* Most of the mixtures designed were coarse-graded Superpave mixes, which are typical of the mixtures being designed and constructed in the State of Arkansas. Gradations for the 16 mix designs are given in Tables 3 - 6. All mixes contained the same grade of binder (PG 64-22).

SS Source	Mix Gradations (Percent Passing)			
	9.5mm	12.5mm	25.0mm	37.5mm
2”	100.0	100.0	100.0	100.0
1-1/2”	100.0	100.0	100.0	100.0
1”	100.0	100.0	93.4	90.6
3/4"	100.0	99.9	81.3	73.3
1/2”	100.0	94.2	66.8	54.7
3/8”	96.4	86.0	58.2	46.2
#4	67.2	56.9	37.2	28.6
#8	37.2	32.3	20.7	16.6
#16	25.5	22.4	15.0	12.2
#30	20.2	17.8	12.5	10.4
#50	16.9	15.0	11.0	9.3
#100	11.3	10.1	7.8	6.7
#200	6.3	5.7	4.8	4.2

Table 3. Blend Gradations for Sandstone Aggregate Source.

SY Source	Mix Gradations (Percent Passing)			
	9.5mm	12.5mm	25.0mm	37.5mm
2”	100.0	100.0	100.0	100.0
1-1/2”	100.0	100.0	100.0	100.0
1”	100.0	100.0	97.5	97.7
3/4"	100.0	99.8	83.7	85.1
1/2”	99.9	91.3	61.9	64.6
3/8”	95.2	81.2	50.4	53.3
#4	71.7	58.1	30.4	32.5
#8	46.8	38.6	21.1	21.8
#16	30.8	25.9	15.5	15.2
#30	20.7	17.7	11.7	11.0
#50	12.7	11.0	7.5	6.9
#100	7.1	6.1	4.6	4.1
#200	4.1	3.5	2.9	2.5

Table 4. Blend Gradations for Syenite Aggregate Source.

GR Source	Mix Gradations (Percent Passing)			
	9.5mm	12.5mm	25.0mm	37.5mm
2”	100.0	100.0	100.0	100.0
1-1/2”	100.0	100.0	100.0	100.0
1”	100.0	100.0	98.5	97.5
3/4"	100.0	99.9	93.3	89.4
1/2”	99.7	94.6	69.4	64.2
3/8”	94.0	84.3	47.4	43.6
#4	62.9	53.9	30.7	29.6
#8	41.8	34.5	21.8	21.1
#16	30.6	24.6	16.5	15.1
#30	22.9	18.0	12.8	11.3
#50	14.7	11.5	8.7	7.4
#100	8.6	6.7	5.4	4.5
#200	5.2	4.1	3.5	3.0

Table 5. Blend Gradations for Gravel Aggregate Source.

LS Source	Mix Gradations (Percent Passing)			
	9.5mm	12.5mm	25.0mm	37.5mm
2”	100.0	100.0	100.0	100.0
1-1/2”	100.0	100.0	100.0	100.0
1”	100.0	100.0	93.7	90.6
3/4"	100.0	99.9	84.9	77.4
1/2”	99.9	94.8	62.6	51.7
3/8”	93.0	84.3	50.0	40.7
#4	58.0	48.2	31.0	24.3
#8	40.5	33.4	23.8	18.7
#16	27.6	22.5	17.2	13.3
#30	18.5	14.7	13.0	9.5
#50	11.7	8.9	9.7	6.8
#100	7.5	5.5	7.3	5.0
#200	5.6	4.0	5.5	3.7

Table 6. Blend Gradations for Limestone Aggregate Source.

For the Gmm study, triplicate samples were produced for each combination of aggregate size and type. The samples were tested according to three methods – the traditional method (AASHTO T-209) and two newly developed methods (CoreLok and Kuss methods).

For the Gmb study, compactive effort was included as an additional factor. For each of sixteen (16) mix designs produced, triplicate samples were compacted to each of three levels of compaction (high, medium, and low). The number of gyrations corresponding with these compactive efforts was estimated for each mix design, corresponding with target air void levels of approximately 2.0 percent, 4.5 percent, and 7.0 percent. This spread represents a range of typical values found in both laboratory- and field-compacted HMA samples. Thus, 144 compacted HMA samples were prepared for the bulk specific gravity analysis. Five test methods were performed in the study including two traditional methods (the SSD and Height-Diameter methods), and three newer methods (CoreLok, CoreReader, and Kuss methods). When required, the paraffin-coating method was performed according to AASHTO T-275. Additional compacted HMA samples from previous research project were used in order to validate the conclusions of the study.

TEST RESULTS AND ANALYSIS

A comprehensive discussion of the statistical analyses is presented in the following sections of this report. SAS statistical software was used to complete the analyses. A five percent level of significance (alpha = 0.05) was used in all cases.

Gmm Methods

A total of 48 samples were tested according to three methods for measuring the maximum theoretical specific gravity of the HMA mix. These samples represented four aggregate sources and four levels of NMAS. Three samples were prepared for each combination of factors, then tested according to each method. A summary of factors and levels is presented in Table 7.

Factor	**# of Levels**	**Levels**
Source	4	Limestone (LS), Sandstone (SS), Gravel (GR), Syenite (SY)
NMAS	4	9.5mm, 12.5mm, 25.0mm, 37.5mm
Gmm Method	3	AASHTO T-209 CoreLok Kuss Method (KS)

Table 7. Summary of ANOVA Factors for Gmm Method Analysis

Analysis of Variance (ANOVA) was used to statistically compare the effect of test method, while also considering the effects of NMAS and aggregate source. Since aggregate source affects density, it was considered even though it was not the variable of interest and had no practical bearing on factor interactions. A complete randomized block design was used to determine if the main effects of Method, NMAS, or the interaction of the two displayed significance. A summary of results is given in Table 8, including the degrees of freedom, calculated F-statistic, and P-value for each parameter. The P-value is the smallest level of significance at which the data are significant. In other words, if the P-value is less than alpha (0.05), then the factor or interaction is significant.

Factor	df	F-calc	P-value
Method	2	99.97	<0.0001
NMAS	3	20.07	<0.0001
Method*NMAS	6	6.45	<0.0001
Source	3	94.79	<0.0001
Error	123	MSE = 0.003877	

Table 8. ANOVA Results for Gmm Method with Interaction

All factors and interactions were significant. Source had a significant effect, meaning that it was beneficial to separate the significant amount of variability created by that factor. Method and NMAS were both significant factors. However, because the interaction of Method and NMAS was significant, the main effects were not considered individually. A significant interaction means that the conclusions for one factor are dependant on another factor, and can be seen as non-parallel lines on an interaction plot. When a significant interaction exists, the characteristics of the interaction should be considered rather than the main effects. In this case, (see Figure 5) the 12.5mm plot appeared to be different from the others. However, there didn't seem to be a pattern with respect to increasing or decreasing NMAS. Therefore, this plot revealed no practically significant pattern or trend.

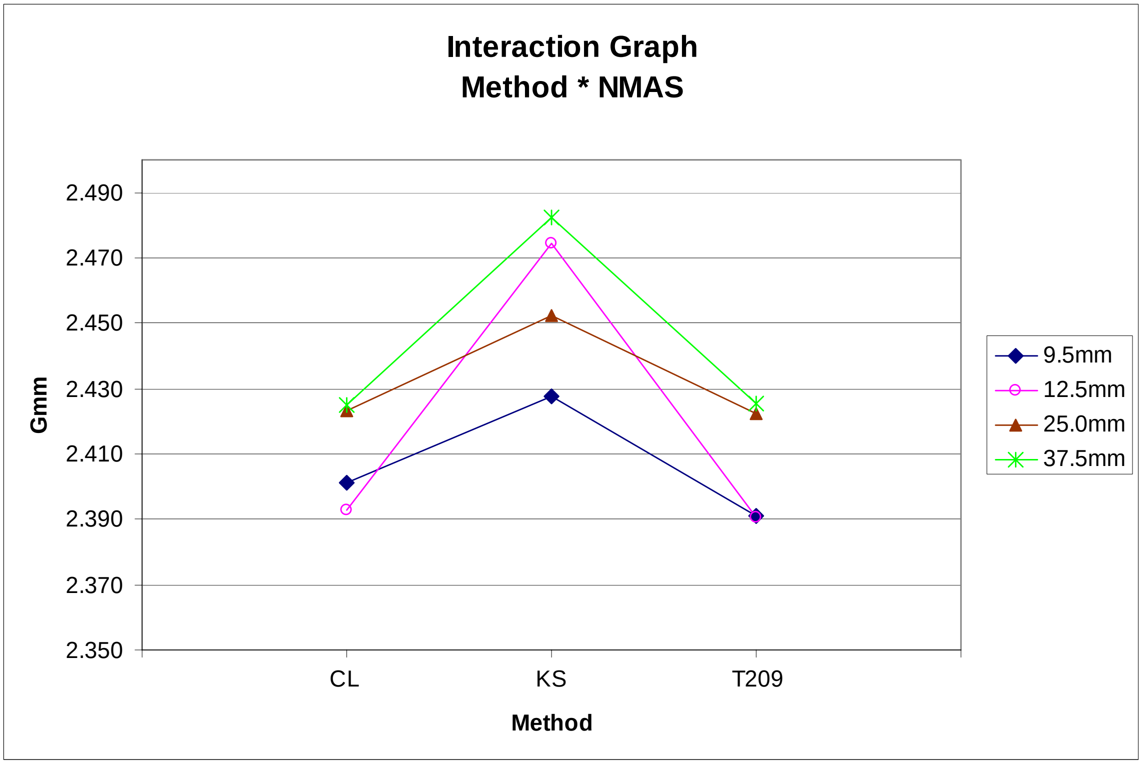

Figure 5. Gmm Interaction Graph – Method * NMAS

Because the interaction plot failed to provide meaningful conclusions, the ANOVA was repeated without the interaction term. These results are given in Tables 9 and 10.

Factor	df	F-calc	P-value
Method	2	81.09	<0.0001
NMAS	3	16.13	<0.0001
Source	3	75.71	<0.0001
Error	137	MSE = 0.06269	

Table 9. ANOVA Results for Gmm Method without Interaction

Means Test for Method				Means Test for NMAS		
Method	Mean	Rank		NMAS	Mean	Rank
Kuss	2.4598	A		37.5mm	2.4453	A
CoreLok	2.4097	B		25.0mm	2.4329	B
T-209	2.4076	B		12.5mm	2.4192	C
				9.5mm	2.4105	C

Table 10. Duncan's Test Results for Gmm Method without Interaction

Method and NMAS were both significant factors. Duncan's Multiple Range Test was used to determine which means caused the significance. Means with the same letter grouping were ranked similarly. In terms of method, a means test reveals that the traditional (T-209) and the CoreLok methods provided similar results, but results provided by the Kuss method were significantly higher. In terms of NMAS, the 9.5mm and 12.5mm mixes were ranked similarly, but the 25.0mm and 37.5mm were each ranked differently.

To further consider the data, a separate ANOVA was performed on each combination of aggregate source and NMAS. In essence, this removed all effects of the aggregate so that the focus could be placed on differences in test method. Although this analysis provided a more intuitive description of the data, it was not as robust because fewer data points were available for each analysis. A single factor ANOVA was used to determine the effect of method on each mix, and if a significant difference existed, Duncan's Multiple Range Test was used to determine which method(s) caused the difference. A summary of results is given in Table 11. For each mix, methods having the same letter ranking showed no statistically significant difference.

			Duncan's Multiple Range Test Results		
Source	NMAS	Mix ID	AASHTO T-209	CoreLok	Kuss Method
SS	9.5mm	A9	A	A	A
SS	12.5mm	A12	A	A	B
SS	25.0mm	A25	A	A	A
SS	37.5mm	A37	A	A	B
SY	9.5mm	G9	A	A	A
SY	12.5mm	G12	A	A	B
SY	25.0mm	G25	A	A	B
SY	37.5mm	G37	A	A	B
GR	9.5mm	J9	A	B	C
GR	12.5mm	J12	A	A	B
GR	25.0mm	J25	A	A	A
GR	37.5mm	J37	A	AB	B
LS	9.5mm	M9	A	A	B
LS	12.5mm	M12	A	A	B
LS	25.0mm	M25	A	A	B
LS	37.5mm	M37	A	A	B

Table 11. ANOVA Results for Gmm Method Analysis by Mix

Even when separated by mix, the T-209 and CoreLok methods almost always exhibited no statistically significant difference, while the Kuss method generated a significantly higher Gmm than the other two. From a practical standpoint, this conclusion was true as well. Table 12 indicates a small average difference in the T-209 and CoreLok methods (0.004). No significant trends were noted regarding aggregate size or type.

Gmm Variability

In order to determine a relative measure of the variability of each method, standard deviation and coefficient of variation were calculated for each mixture. A summary of data is contained in Table 12.

			AASHTO T-209			CoreLok			Kuss Method		
Source	NMAS	Mix ID	Avg. Gmm	Std. Dev.	COV (%)	Avg. Gmm	Std. Dev.	COV (%)	Avg. Gmm	Std. Dev.	COV (%)
SS	9.5mm	A9	2.379	0.00361	0.152	2.383	0.00289	0.121	2.386	0.02079	0.871
SS	12.5mm	A12	2.362	0.00321	0.136	2.368	0.00436	0.184	2.430	0.01050	0.432
SS	25.0mm	A25	2.395	0.00321	0.134	2.405	0.00265	0.110	2.402	0.01201	0.500
SS	37.5mm	A37	2.397	0.00200	0.083	2.404	0.00862	0.359	2.423	0.01069	0.441
SY	9.5mm	G9	2.428	0.00424	0.175	2.423	0.02263	0.934	2.452	0.01414	0.577
SY	12.5mm	G12	2.421	0.00416	0.172	2.418	0.00462	0.191	2.509	0.01115	0.444
SY	25.0mm	G25	2.470	0.00577	0.234	2.469	0.00872	0.353	2.533	0.03204	1.265
SY	37.5mm	G37	2.467	0.00451	0.183	2.475	0.01556	0.629	2.569	0.00058	0.022
GR	9.5mm	J9	2.358	0.00361	0.153	2.395	0.00153	0.064	2.415	0.01706	0.706
GR	12.5mm	J12	2.391	0.00115	0.048	2.391	0.01249	0.522	2.461	0.00603	0.245
GR	25.0mm	J25	2.410	0.00751	0.311	2.410	0.00252	0.104	2.413	0.01872	0.776
GR	37.5mm	J37	2.420	0.00321	0.133	2.429	0.00424	0.175	2.439	0.00800	0.328
LS	9.5mm	M9	2.411	0.00252	0.104	2.410	0.00153	0.063	2.465	0.01677	0.681
LS	12.5mm	M12	2.389	0.00551	0.231	2.393	0.01514	0.633	2.497	0.00907	0.363
LS	25.0mm	M25	2.414	0.00416	0.172	2.403	0.02051	0.854	2.461	0.02951	1.199
LS	37.5mm	M37	2.418	0.01058	0.438	2.410	0.00351	0.146	2.498	0.00872	0.349
Average Values			2.408	0.00431	0.179	2.412	0.00822	0.340	2.460	0.01411	0.575

Table 12. Summary of Statistics for Gmm Values

From the table, it is evident that the traditional method (AASHTO T-209) exhibited the lowest level of variability, having a COV of just 0.179 percent. The CoreLok had almost twice the variability of T-209, and the Kuss method had just over three times the variability of T-209. However, all three methods demonstrated relatively low levels of variability, having COV values less than 1 percent.

Gmm Discussion

Since there is no way to measure an absolutely "true" value of Gmm, it should not be said that the Kuss method produced inaccurate results. It did, however, produce results that were significantly higher than the other methods, in both a practical sense and a statistical sense. The most likely reason for this difference is that the Kuss method used air to permeate the sample rather than water. Air can flow more easily into the smallest voids of the sample, producing a truer measure of apparent volume. This provided a smaller sample volume and thus a greater sample density. While this method may provide a more accurate measure of density, the fact that it produced "different" values means that current design procedures would have to be changed in order to accommodate this difference. The procedure could be calibrated to aid in the adjustment, but the relatively large variability does not make this an attractive option. A significant advantage such as lower variability would be necessary to warrant the changes necessary to incorporate this test method into current specifications.

In terms of variability, the T-209 method was best, although all three methods could be termed "good" and exhibited low levels of variability. Based on these results, it was concluded that the traditional method (AASHTO T-209) is the preferred method for the measurement of maximum theoretical specific gravity of asphalt mixes, although the CoreLok method also showed promise. While the accuracy of the test methods could only be assessed relatively, the T-209 and CoreLok methods did provide similar results. In terms of variability, the T-209 method was more repeatable than the CoreLok method. If the CoreLok method is to be used, steps should be taken to increase the precision of the method. Because the Kuss method had the largest variability, it should not be used for routing testing until modifications are made. Considering ease of use and time required for testing, the Kuss method would be a preferred method due to the quick return of test results, and should be studied further. The CoreLok method was somewhat faster than T-209 to perform, requiring approximately half the time of the T-209 method. The most time consuming part of running these tests was sample preparation, which involved separating individual particles of the mixture. This step had to be completed for all three methods, so there was no real advantage in this regard. Unless the variability of an alternative method is improved, there is not enough evidence to recommend the addition of a new test method for maximum theoretical specific gravity to current specifications at the present time.

Gmb Methods

The first focus of this analysis was to compare the density measurements generated by the various test methods. In order to do this, the data was first plotted in several different manners to see what noticeable trends were present. Plots demonstrating the Gmb values as measured by the five methods are presented in Figures 7 – 10. Each Figure contains data for 48 samples corresponding to a single NMAS, and is sorted according to aggregate source and level of compactive effort.

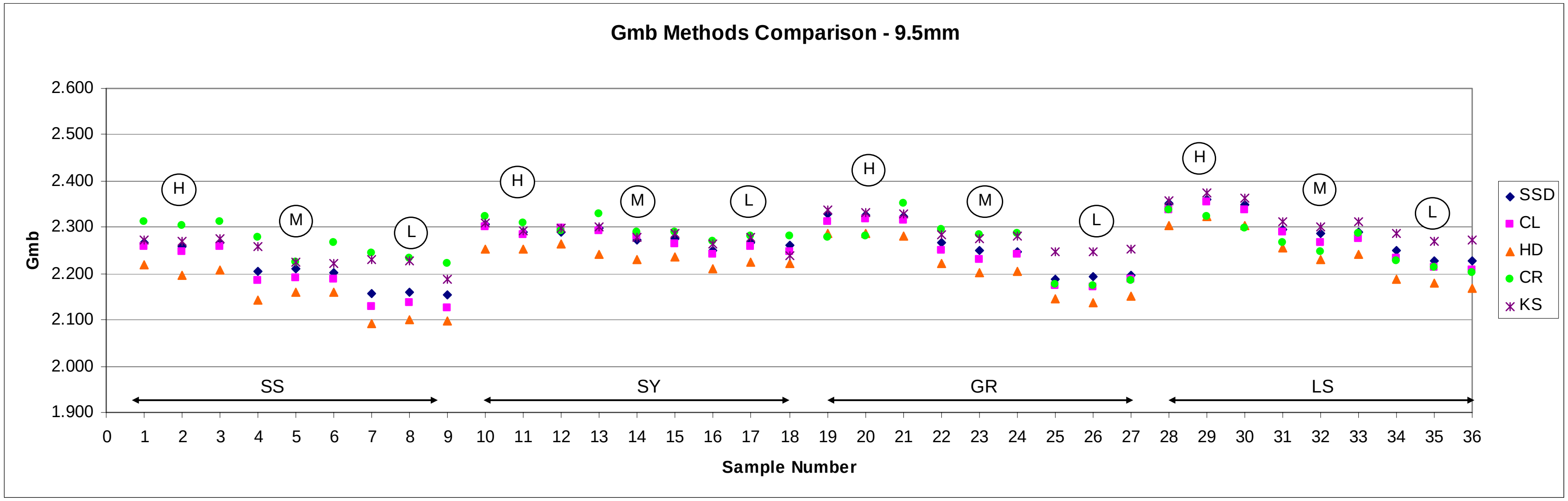

Figure 7. Comparison of five methods measuring Gmb for mixes with NMAS = 9.5mm

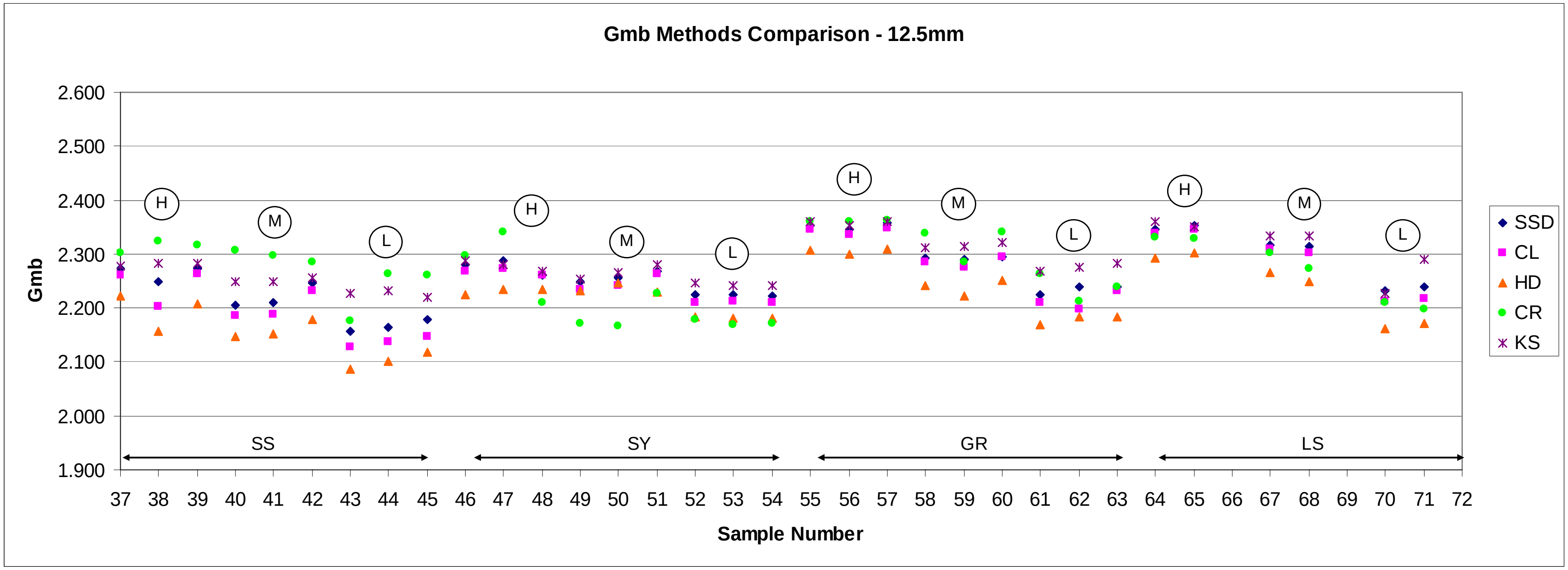

Figure 8. Comparison of five methods measuring Gmb for mixes with NMAS = 12.5mm

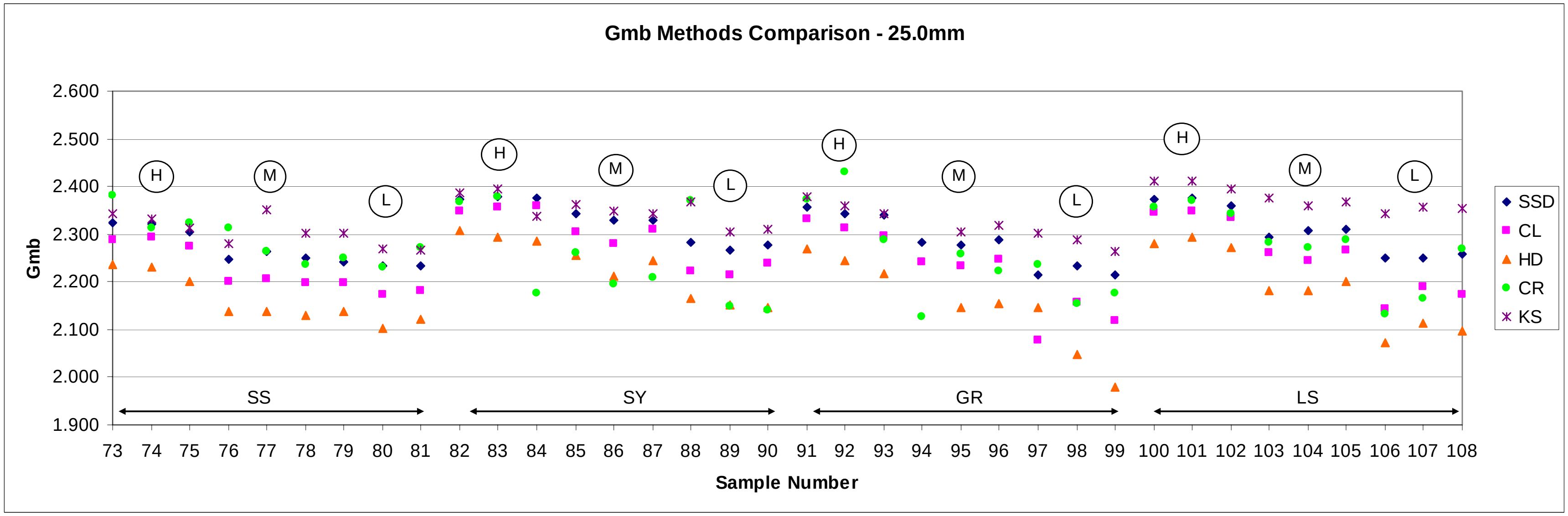

Figure 9. Comparison of five methods measuring Gmb for mixes with NMAS = 25.0mm

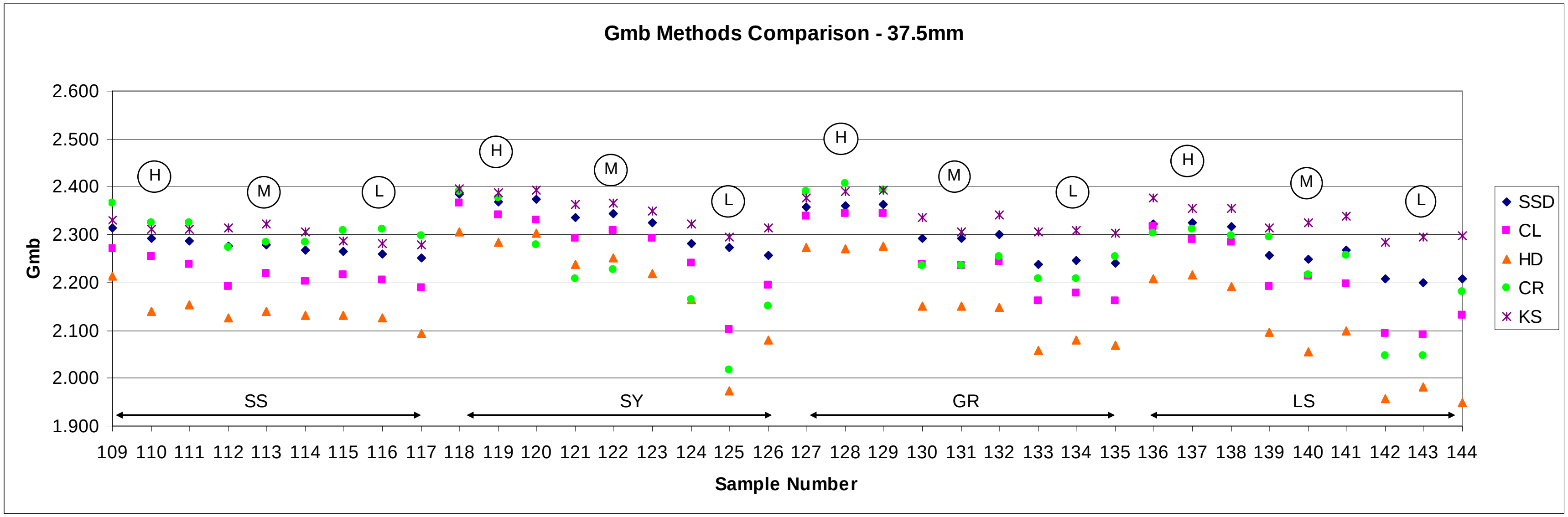

Figure 10. Comparison of five methods measuring Gmb for mixes with NMAS = 37.5mm

By visual inspection, the sample groupings were evident, densities decreasing as compactive effort decreased. Two important trends presented themselves. First, as the nominal maximum size of the aggregate increased, the spread of Gmb values among the different methods also increased. In other words, the average distance between the minimum and maximum data points was larger for the 25.0mm and 37.5mm NMAS mixes than for the 9.5mm and 12.5mm mixes. This is reasonable since specimens with a larger NMAS are prone to greater surface irregularities, which has been demonstrated to create greater discrepancies among measurement methods.

Secondly, specimens prepared with lesser compactive efforts seemed to display increased variability. Again, this can be expected because smaller compactive efforts allow for greater air void contents, increasing the likelihood of interconnected void pathways and greater surface irregularities.

In order to quantify these trends, ANOVA procedures were used along with means tests (Duncan's Multiple Range Test and Least Squares Means Tests) to determine which factors significantly affected test results, and which levels of factors caused the significance, when differences existed. Close attention to the "spread" of data for each sample in Figures # - # indicated a practically significant difference between methods. The ANOVA procedures were used to attempt to quantify this conclusion.

In the first analysis, a complete randomized block design was used to test the effects of four factors for potential effects on Gmb measurements. These factors are summarized in Table 13. Source was treated as a block since different aggregates have different densities, and add variability to the Gmb values. The Source factor added variability which could not be controlled, yet presented no practical interacting effects.

Factor	# of Levels	Levels
Source	4	Limestone (LS), Sandstone (SS), Gravel (GR), Syenite (SY)
NMAS	4	9.5mm, 12.5mm, 25.0mm, 37.5mm
Compactive Effort	3	Low, Medium, High
Gmb Method	5	AASHTO T-166 (SSD) CoreLok (CL) Height-Diameter (HD) CoreReader (CR) Kuss Method (KS)

Table 13. Summary of ANOVA Factors

The results of this analysis, summarized in Table 14, indicated that source was a significant factor, so treating it as a block was beneficial in removing “noise” from the data. The three-way interaction (NMAS * Compactive Effort * Method) was insignificant, but all two-way interactions and main effects were significant.

Factor	df	F-value	P-value
Method	4	177.54	<0.0001
NMAS	3	4.43	0.0043
CompEff	2	435.94	<0.0001
Method*NMAS	12	11.22	<0.0001
Method*CompEff	8	6.22	<0.0001
CompEff*NMAS	6	3.96	0.0007
Method*NMAS*CompEff	24	0.94	0.5450
Source	3	31.53	<0.0001
Error	639	MSE = 0.001597	

Table 14. ANOVA Summary for Gmb Values

A significant interaction was detected for the Method and NMAS factors. In other words, trends for one factor (Method) were somewhat dependent on another factor

(Compactive Effort). In this case (shown in Figure 11), the Height-Diameter method was most sensitive to changes in NMAS, and the CoreLok and CoreReader methods appeared to be the least affected by NMAS.

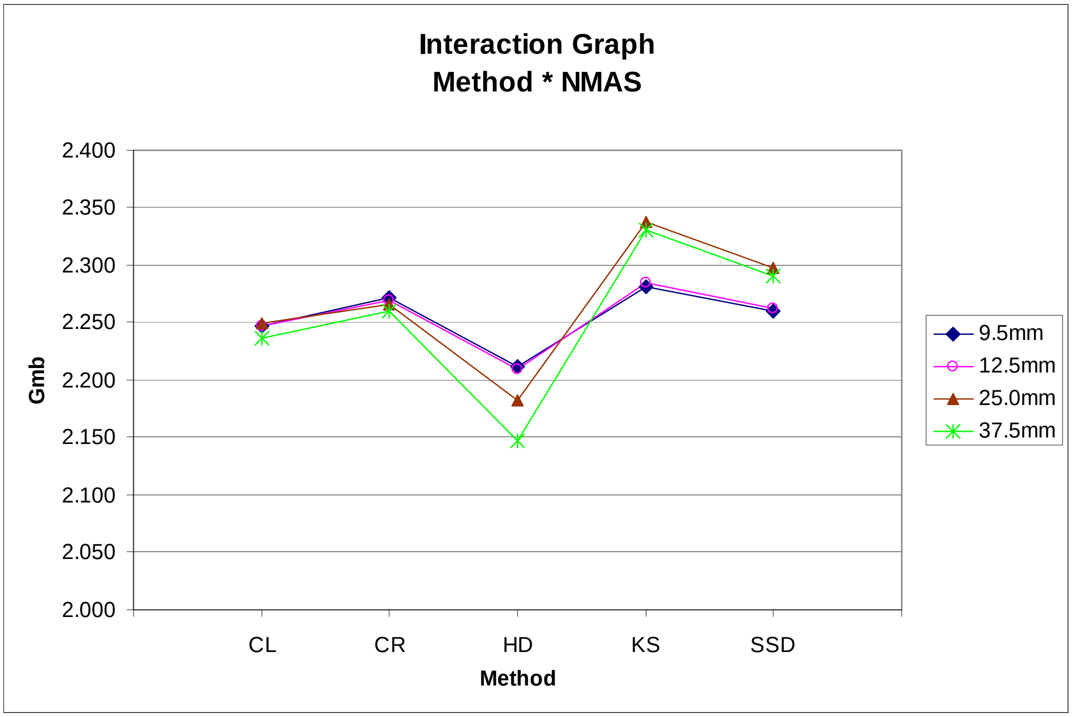

Figure 11. Interaction Graph – Method * NMAS

A significant interaction was also detected for Method and Compactive Effort (shown in Figure 12). The primary conclusion in this case was that Gmb values measured by the Kuss method appeared to be the least affected by variations in compactive effort.

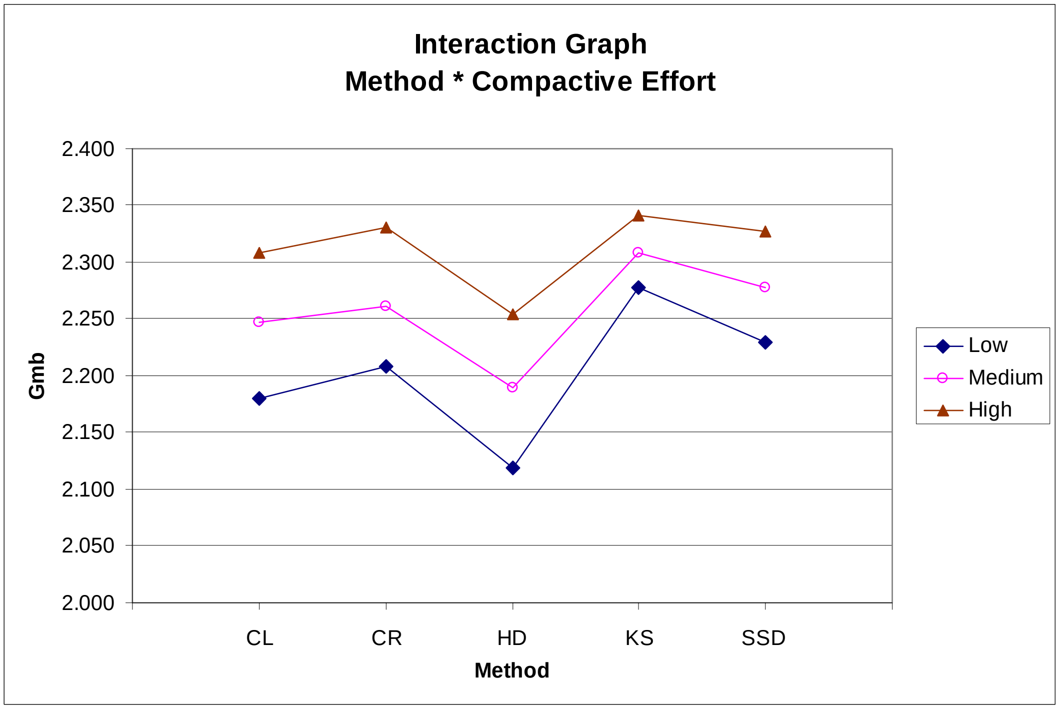

Figure 12. Interaction Graph – Method * Compactive Effort

Finally, a significant interaction was detected for the factors of Compactive Effort and NMAS (shown in Figure 13). The most prominent trend noted here is that average Gmb values for the smaller aggregate sizes seemed to be less affected by changes in compactive effort. Mixes with smaller aggregates are less prone to interconnecting void pathways, so this conclusion is reasonable.

Although it is not proper to consider the main effects when significant interactions involving those main effects are present, it is noted that the means test ranked each method separately, indicating that all methods produced statistically different results.

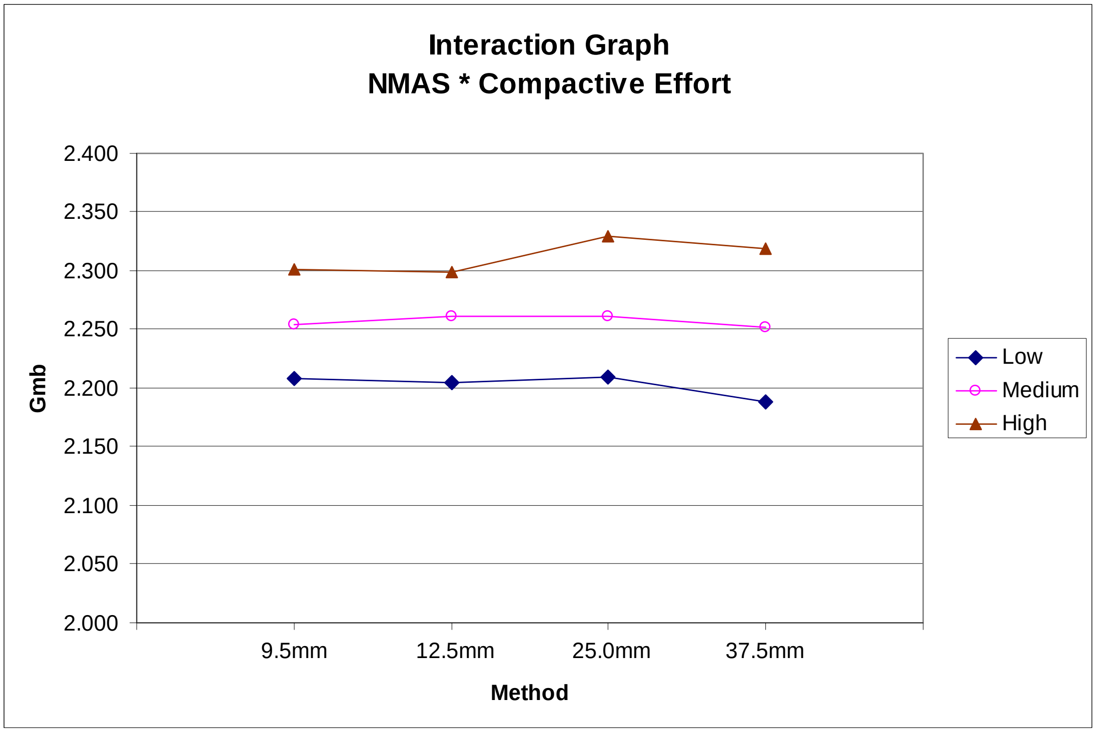

Figure 13. Interaction Graph – NMAS * Compactive Effort

In order to more intuitively examine the data, separate analyses were conducted for each mix. Though the number of data points in each analysis was significantly reduced, the effects and interactions of Source, NMAS, and Compactive Effort could essentially be removed from the data, and the effect of test method could be analyzed separately. The results for the 48 cases, including the mean and rank for each method, are summarized in Table 15.

Although analyzing the data in this way is not as robust and does not allow for the statistical detection of significant trends for factors such as NMAS and level of compactive effort, it does provide an idea of "how different" the methods are for the various mixture types.

			Mean and Rank According to ANOVA and Duncan's Multiple Range Test									
			CoreLok		CoreReader		Height-Diameter		Kuss		SSD	
Source	NMAS	Comp. Effort	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank
SS	9.5	High	2.255	C	2.308	A	2.208	D	2.272	B	2.264	BC
SS	9.5	Medium	2.188	B	2.256	A	2.154	C	2.235	A	2.205	B
SS	9.5	Low	2.130	C	2.233	A	2.097	D	2.215	A	2.157	B
SS	12.5	High	2.242	B	2.314	A	2.195	C	2.281	AB	2.266	B
SS	12.5	Medium	2.201	C	2.297	A	2.160	D	2.251	B	2.220	BC
SS	12.5	Low	2.137	BC	2.233	A	2.102	C	2.226	A	2.167	B
SS	25.0	High	2.285	B	2.340	A	2.223	C	2.329	A	2.317	AB
SS	25.0	Medium	2.202	C	2.272	AB	2.135	D	2.311	A	2.254	B
SS	25.0	Low	2.184	C	2.251	AB	2.120	D	2.280	A	2.237	B
SS	37.5	High	2.254	B	2.339	A	2.169	C	2.318	A	2.297	A
SS	37.5	Medium	2.205	C	2.280	B	2.133	D	2.314	A	2.274	B
SS	37.5	Low	2.203	D	2.306	A	2.118	E	2.282	B	2.258	C
SY	9.5	High	2.294	A	2.308	A	2.256	B	2.299	A	2.295	A
SY	9.5	Medium	2.279	A	2.303	A	2.236	B	2.289	A	2.282	A
SY	9.5	Low	2.248	B	2.277	A	2.219	C	2.260	AB	2.260	AB
SY	12.5	High	2.237	A	2.284	A	2.231	A	2.279	A	2.276	A
SY	12.5	Medium	2.246	A	2.189	B	2.235	A	2.267	A	2.258	A
SY	12.5	Low	2.211	C	2.173	E	2.182	D	2.243	A	2.224	B
SY	25.0	High	2.355	A	2.307	A	2.295	A	2.372	A	2.376	A
SY	25.0	Medium	2.299	B	2.222	C	2.237	C	2.350	A	2.334	AB
SY	25.0	Low	2.225	AB	2.220	AB	2.155	B	2.328	A	2.275	A
SY	37.5	High	2.345	AB	2.348	AB	2.298	B	2.391	A	2.376	A
SY	37.5	Medium	2.298	C	2.218	D	2.235	D	2.359	A	2.334	B
SY	37.5	Low	2.178	BC	2.111	C	2.072	C	2.310	A	2.270	AB
GR	9.5	High	2.314	AB	2.303	AB	2.285	B	2.333	A	2.327	A
GR	9.5	Medium	2.241	B	2.289	A	2.209	C	2.280	A	2.255	B
GR	9.5	Low	2.178	C	2.179	C	2.146	D	2.250	A	2.193	B
GR	12.5	High	2.343	B	2.361	A	2.306	C	2.358	A	2.352	A
GR	12.5	Medium	2.285	B	2.322	A	2.239	C	2.316	AB	2.292	AB
GR	12.5	Low	2.213	B	2.238	B	2.179	C	2.275	A	2.234	B
GR	25.0	High	2.315	A	2.364	A	2.244	B	2.360	A	2.346	A
GR	25.0	Medium	2.241	AB	2.203	BC	2.150	C	2.311	A	2.282	A
GR	25.0	Low	2.118	CD	2.189	BC	2.057	D	2.285	A	2.221	AB
GR	37.5	High	2.342	C	2.396	A	2.274	D	2.386	A	2.360	B
GR	37.5	Medium	2.238	C	2.240	C	2.150	D	2.327	A	2.295	B
GR	37.5	Low	2.167	C	2.222	B	2.069	D	2.306	A	2.241	B
LS	9.5	High	2.343	A	2.319	B	2.310	B	2.364	A	2.353	A
LS	9.5	Medium	2.278	B	2.267	B	2.242	C	2.309	A	2.290	AB
LS	9.5	Low	2.219	B	2.214	B	2.179	C	2.276	A	2.234	B
LS	12.5	High	2.343	AB	2.330	B	2.297	C	2.356	A	2.349	A
LS	12.5	Medium	2.306	AB	2.288	B	2.258	C	2.334	A	2.316	AB
LS	12.5	Low	2.216	AB	2.205	AB	2.167	B	2.259	A	2.235	A
LS	25.0	High	2.343	C	2.357	BC	2.283	D	2.406	A	2.369	B
LS	25.0	Medium	2.257	D	2.280	C	2.188	E	2.348	A	2.304	B
LS	25.0	Low	2.169	C	2.189	C	2.095	D	2.350	A	2.253	B
LS	37.5	High	2.297	C	2.304	BC	2.206	D	2.362	A	2.321	B
LS	37.5	Medium	2.203	C	2.255	B	2.084	D	2.326	A	2.257	B
LS	37.5	Low	2.105	C	2.091	C	1.962	D	2.292	A	2.204	B

Table 15. Summary of ANOVA and Duncan's Test Results for Individual Mixes

In almost all cases (45 of 48), the Kuss method was assigned the letter 'A', meaning that it was in the group that has the largest Gmb values (i.e., higher densities). The SSD method was assigned the letter 'A' or 'B' in all but one case, which means that it produced relatively high density values. The CoreLok method was generally in the middle of the density range, receiving intermediate letter rankings. This is consistent with the cursory conclusions drawn from visual examination of Figures 7 - 10. The CoreReader was a member of several different rankings, and thus seemed to be more variable. In other words, sometimes it produced higher Gmb values, and sometimes it produced lower Gmb values. The Height-Diameter method received the lowest letter ranking in all but one case. This was true even though this method was assigned several different letter codes. To explain, if there were two letter rankings and the Height-Diameter method received a 'B', it is in the lowest letter ranking. On the other hand, if there were five letter rankings and the Height-Diameter method received an 'E', it is still the lowest letter ranking even though the letter names were different. In three cases, all five methods received separate rankings, meaning that all five methods produced results that were different enough to be considered statistically significant. In two cases, all five methods received the same ranking, meaning that for those two mixes, all five methods produced similar results.

The SSD and CoreLok methods were ranked similarly in 27 of 48 cases (56 percent), which is less than that reported by Buchanan, where they were similar in 79 percent of the cases. *(3)* However, Buchanan's study considered only 12.5mm mixes. Considering only surface mixes in this study, the SSD and CoreLok methods were ranked similarly in 20 of 24 cases (83 percent), which is consistent with Buchanan's findings.

Another way to consider the effects of different methods is to look at the magnitude of departure from the traditional, or standard, result. This is important because of the fact that there is no “true” measure of Gmb, and so there is no way to quantify the “error” of a test method. By analyzing the difference in test results, rather than actual Gmb values, a relative sense of accuracy can be evaluated, as well as helping to describe possible impacts that would be associated with changing current specifications to include additional methods. By comparing relative measures of Gmb, the traditional method (SSD) forms a baseline, and a more complete sense of the impact of changing methods and/or specifications can be attained.

In this portion of the analysis, the *difference* in Gmbs as measured by SSD and the other methods was the variable of interest. Since the SSD values were typically higher than the others, these differences were calculated by subtracting the Gmb by the alternative methods from the SSD-measured Gmb. A summary of the factors and levels for this analysis is given in Table 16.

Factor	**# of Levels**	**Levels**
Source	4	Limestone (LS), Sandstone (SS), Gravel (GR), Syenite (SY)
NMAS	4	9.5mm, 12.5mm, 25.0mm, 37.5mm
Compactive Effort	3	Low, Medium, High
Gmb Method	4	SSD – CoreLok (SSD-CL) SSD – CoreReader (SSD-CR) SSD – Kuss Method (SSD-KS) SSD – Height-Diameter (SSD-HD)

Table 16. Summary of ANOVA Factors

The results of this analysis (given in Table 17) indicated that again, source was a significant source of data variability, and the blocking procedure was beneficial in separating this portion of the variability. The three-way interaction (NMAS * Compactive Effort * Method) was not significant, but all two-way interactions and main effects were significant.

Factor	df	F-value	P-value
Method	3	314.58	<0.0001
NMAS	3	50.48	<0.0001
CompEff	2	7.39	0.0007
Method*NMAS	9	17.80	<0.0001
Method*CompEff	6	11.23	<0.0001
CompEff*NMAS	6	5.85	<0.0001
Method*NMAS*CompEff	18	1.37	0.1388
Source	3	10.59	<0.0001
Error	510	MSE = 0.0011265	

Table 17. ANOVA Summary for Gmb Values

A significant interaction was detected for the Method and NMAS factors. In this case (shown in Figure 14), the difference between the SSD and Height-Diameter methods was most sensitive to changes in NMAS, with the greatest differences occurring for the 25.0mm and 37.5mm mixes. The differences in the SSD and Kuss methods appeared to be least affected by changes in NMAS.

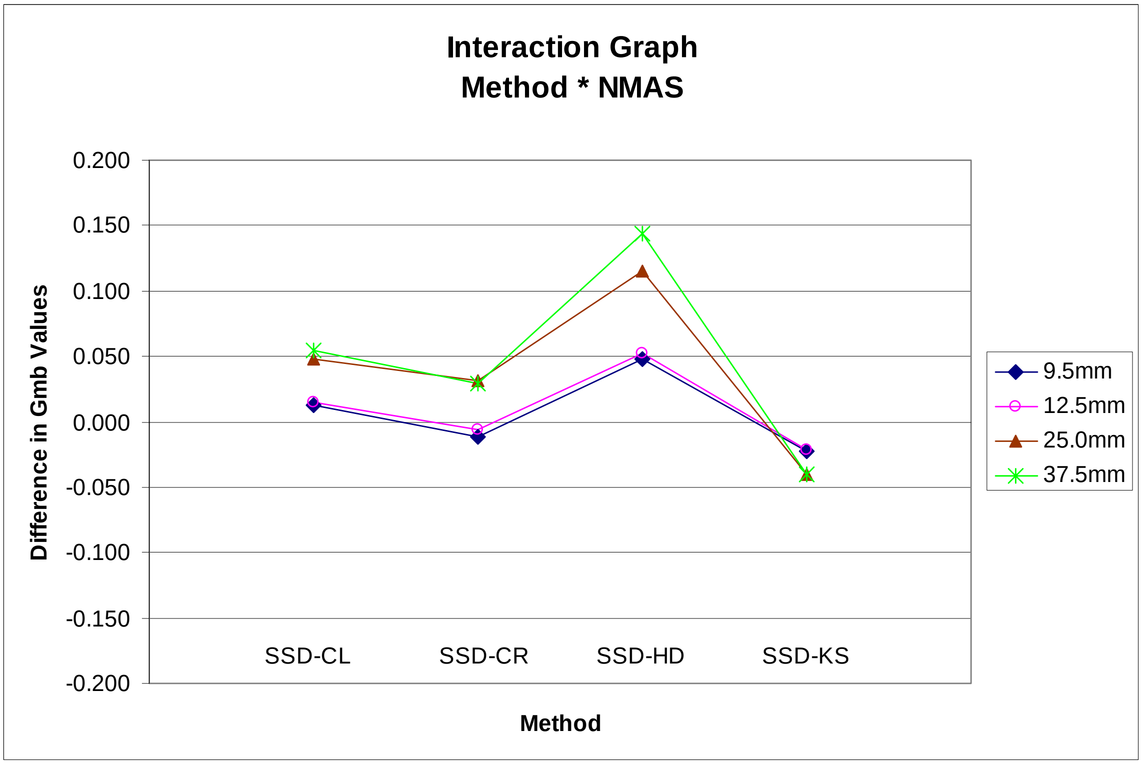

Figure 14. Interaction Graph – Method * NMAS

A significant interaction was also detected for Method and Compactive Effort (shown in Figure 15). The primary reason for this interaction is that the differences in the SSD and Kuss methods were negative, whereas the other differences are positive. Thus the pattern for differences in low, medium and high compactive effort was reversed. Due to this phenomenon, the interaction of Method and Compactive Effort really has no practical significance. In fact, the ANOVA was repeated using the absolute values of the differences in methods, and this interaction was not significant. So when properly considered, the effects of this potential interaction were not significant.

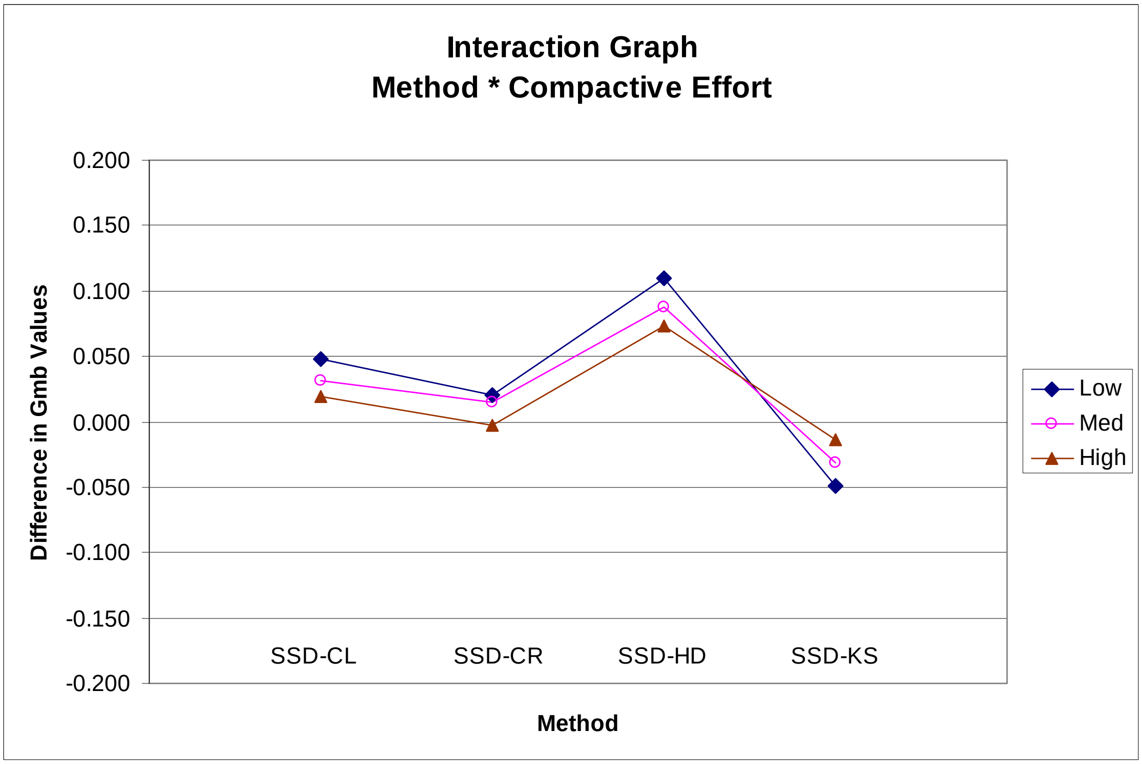

Figure 15. Interaction Graph – Method * Compactive Effort

Finally, a significant interaction was detected for the factors of Compactive Effort and NMAS (shown in Figure 16). The most prominent trend noted here is that the differences in Gmb values for the larger aggregate sizes seemed to be more affected by changes in compactive effort, especially the 37.5mm NMAS. This plot clearly demonstrates the idea that mixes with larger aggregate particles also contain larger air void spaces, which are more likely to be interconnected. These differences were more pronounced for low compactive efforts.

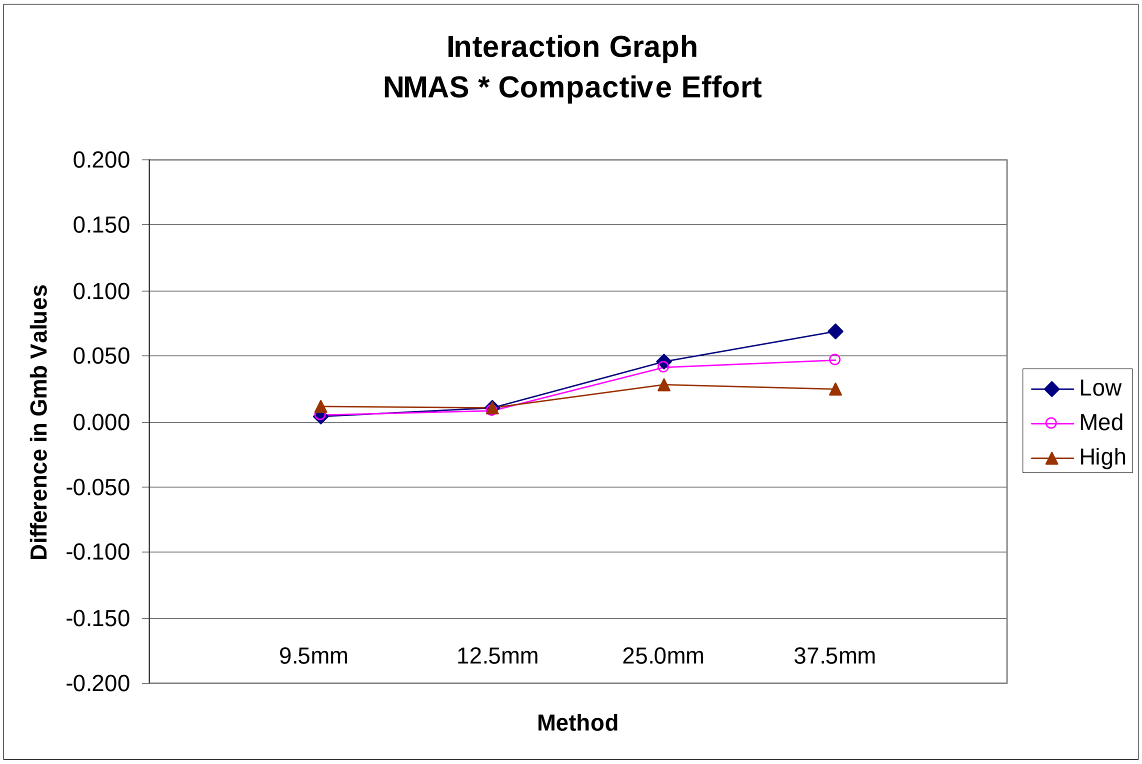

Figure 16. Interaction Graph – NMAS * Compactive Effort

As stated previously, because there were significant interactions within the dataset, main effects should not be considered individually. However, it is interesting to note that the means test ranked the differences by each method in a separate grouping.

In conclusion, the results of several statistical analyses were consistent. Bulk densities of compacted HMA samples were significantly affected by the method of measurement, nominal maximum aggregate size, and level of compactive effort, as well as combinations of these factors. Gmb measurements were less reliable as NMAS increased and as compactive effort decreased. This is reasonable since these circumstances increase surface irregularities of samples, creating the conditions that are believed to be responsible for difficulties in measuring Gmb by any method. It seems clear that the Gmb values generated by the various methods did not provide statistically equivalent results.

Gmb Variability

In choosing a method for the measurement of Gmb, it is desirable to generate precise results as well as accurate results. Since there is no true measure of accuracy for Gmb measurements, then the importance of limiting the variability is even greater. And it is possible to quantify the variability of the test methods. In this study, three samples were compacted for each combination of factors, representing a total of 48 sample types and 144 total samples. Five test methods were performed on each sample, after which the standard deviation and coefficient of variation were calculated for each mix in order to evaluate the variability. It is important to note that the variability exhibited by each mix included both the variability of the separate samples (batching, mixing, and compacting) as well as the variability of the test method itself. The same operator performed each of the three tests in order to remove a potential error source.

Table 18 contains summary statistics describing the variability of the various test methods. In terms of standard deviation and coefficient of variation, the SSD method exhibited the smallest level of variability, followed by the Kuss method, the CoreLok, Height-Diameter, and finally the CoreReader. All methods except the CoreLok had good repeatability, having COV values less than 1.0 percent.

			Standard Deviation and Coefficient of Variation (%)									
			CoreLok		CoreReader		Height-Diameter		Kuss		SSD	
Source	NMAS	Comp. Effort	Std. Dev.	COV%	Std. Dev.	COV %	Std. Dev.	COV%	Std. Dev.	COV%	Std. Dev.	COV%
SS	9.5	High	0.0066	0.291	0.0046	0.200	0.0105	0.474	0.0030	0.132	0.0050	0.220
SS	9.5	Medium	0.0039	0.178	0.0284	1.257	0.0097	0.449	0.0206	0.924	0.0037	0.170
SS	9.5	Low	0.0058	0.274	0.0105	0.472	0.0038	0.181	0.0243	1.098	0.0034	0.158
SS	12.5	High	0.0343	1.532	0.0108	0.467	0.0338	1.542	0.0017	0.076	0.0138	0.611
SS	12.5	Medium	0.0262	1.189	0.0101	0.438	0.0167	0.775	0.0052	0.231	0.0227	1.022
SS	12.5	Low	0.0094	0.439	0.0488	2.185	0.0152	0.724	0.0062	0.281	0.0109	0.503
SS	25.0	High	0.0096	0.418	0.0361	1.545	0.0198	0.892	0.0154	0.661	0.0099	0.429
SS	25.0	Medium	0.0034	0.155	0.0384	1.692	0.0050	0.235	0.0365	1.578	0.0080	0.354
SS	25.0	Low	0.0127	0.582	0.0215	0.956	0.0182	0.856	0.0202	0.887	0.0051	0.228
SS	37.5	High	0.0169	0.750	0.0243	1.038	0.0390	1.798	0.0107	0.461	0.0139	0.603
SS	37.5	Medium	0.0134	0.609	0.0052	0.228	0.0068	0.317	0.0080	0.347	0.0065	0.284
SS	37.5	Low	0.0135	0.612	0.0061	0.264	0.0207	0.976	0.0045	0.198	0.0075	0.332
SY	9.5	High	0.0092	0.402	0.0155	0.672	0.0062	0.275	0.0082	0.356	0.0100	0.437
SY	9.5	Medium	0.0139	0.608	0.0228	0.991	0.0052	0.230	0.0112	0.491	0.0142	0.621
SY	9.5	Low	0.0084	0.372	0.0070	0.307	0.0068	0.305	0.0192	0.850	0.0083	0.365
SY	12.5	High	0.0068	0.302	0.0667	2.919	0.0065	0.290	0.0096	0.422	0.0128	0.562
SY	12.5	Medium	0.0143	0.637	0.0342	1.562	0.0088	0.396	0.0130	0.574	0.0101	0.445
SY	12.5	Low	0.0010	0.045	0.0051	0.236	0.0022	0.103	0.0035	0.154	0.0025	0.112
SY	25.0	High	0.0053	0.226	0.1130	4.898	0.0109	0.475	0.0309	1.301	0.0024	0.101
SY	25.0	Medium	0.0159	0.691	0.0346	1.557	0.0232	1.035	0.0097	0.413	0.0067	0.288
SY	25.0	Low	0.0121	0.543	0.1302	5.868	0.0100	0.464	0.0347	1.490	0.0085	0.372
SY	37.5	High	0.0178	0.760	0.0614	2.613	0.0116	0.506	0.0040	0.169	0.0085	0.359
SY	37.5	Medium	0.0085	0.369	0.1930	8.288	0.0160	0.718	0.0093	0.394	0.0098	0.421
SY	37.5	Low	0.0697	3.198	0.0814	3.856	0.0961	4.636	0.0132	0.571	0.0116	0.510
GR	9.5	High	0.0024	0.103	0.0414	1.795	0.0034	0.148	0.0050	0.216	0.0032	0.136
GR	9.5	Medium	0.0110	0.493	0.0047	0.206	0.0111	0.500	0.0045	0.198	0.0115	0.512
GR	9.5	Low	0.0089	0.407	0.0061	0.280	0.0069	0.323	0.0035	0.154	0.0045	0.206
GR	12.5	High	0.0059	0.253	0.0020	0.085	0.0047	0.203	0.0042	0.177	0.0060	0.255
GR	12.5	Medium	0.0106	0.465	0.0315	1.356	0.0144	0.641	0.0046	0.198	0.0024	0.106
GR	12.5	Low	0.0182	0.822	0.0255	1.139	0.0088	0.403	0.0075	0.332	0.0085	0.380
GR	25.0	High	0.0175	0.758	0.0709	3.000	0.0253	1.128	0.0175	0.742	0.0095	0.404
GR	25.0	Medium	0.0069	0.306	0.0685	3.110	0.0069	0.321	0.0092	0.398	0.0062	0.272
GR	25.0	Low	0.0385	1.820	0.0433	1.977	0.0832	4.046	0.0191	0.838	0.0120	0.540
GR	37.5	High	0.0031	0.134	0.0084	0.350	0.0020	0.090	0.0087	0.365	0.0028	0.121
GR	37.5	Medium	0.0041	0.183	0.0110	0.490	0.0015	0.070	0.0197	0.845	0.0045	0.197
GR	37.5	Low	0.0090	0.413	0.0266	1.195	0.0108	0.522	0.0026	0.115	0.0039	0.173
LS	9.5	High	0.0091	0.390	0.0207	0.891	0.0109	0.474	0.0087	0.370	0.0056	0.239
LS	9.5	Medium	0.0119	0.523	0.0200	0.882	0.0135	0.600	0.0059	0.254	0.0049	0.214
LS	9.5	Low	0.0128	0.577	0.0131	0.592	0.0106	0.486	0.0087	0.383	0.0129	0.576
LS	12.5	High	0.0051	0.219	0.0021	0.091	0.0062	0.270	0.0071	0.300	0.0045	0.192
LS	12.5	Medium	0.0043	0.188	0.0198	0.865	0.0124	0.547	0.0007	0.030	0.0011	0.049
LS	12.5	Low	0.0039	0.176	0.0085	0.385	0.0080	0.371	0.0460	2.035	0.0052	0.231
LS	25.0	High	0.0073	0.311	0.0130	0.552	0.0113	0.495	0.0093	0.386	0.0077	0.325
LS	25.0	Medium	0.0113	0.500	0.0082	0.359	0.0115	0.526	0.0075	0.317	0.0081	0.350
LS	25.0	Low	0.0236	1.090	0.0708	3.234	0.0197	0.939	0.0074	0.314	0.0041	0.182
LS	37.5	High	0.0172	0.747	0.0067	0.289	0.0117	0.531	0.0124	0.526	0.0031	0.132
LS	37.5	Medium	0.0113	0.513	0.0390	1.729	0.0243	1.164	0.0120	0.516	0.0094	0.417
LS	37.5	Low	0.0219	1.038	0.0777	3.713	0.0168	0.858	0.0078	0.339	0.0049	0.221
AVERAGE VALUES			0.0128	0.575	0.0344	1.522	0.0154	0.715	0.0117	0.508	0.0075	0.332

Table 18. Standard Deviation and Coefficient of Variation (%) for Individual Mixes

Gmb Discussion

In terms of accuracy, it is difficult to presume which method(s) provide more accurate measures of Gmb. It seems logical that the SSD method probably does underestimate sample volume due to the interconnected void pathways present in the coarser, less compacted mixes. It is also logical that the height-diameter method is likely to overestimate sample volume by treating the sample as a perfectly smooth cylinder. The CoreLok measures of Gmb in this research have proven to be moderate values – greater than that of the height-diameter method and less than that of the SSD method. A "middle of the road" value seems to be a safe bet when no absolute answer is available. Also, no major logical flaws in the concept of the method readily presented themselves.

In light of this, the CoreLok could be a viable alternative for measuring Gmb, but the effects of the different values must first be assessed in order to justify changing the existing specifications.

As for variability, the SSD method was the least variable method, followed by the CoreLok and Kuss methods. The data presented in this study simply does not support the elimination of the SSD method for measuring bulk specific gravity of compacted HMA samples.

Additional Topics

The Paraffin Method

When a sample is tested by the SSD method and has greater than 2.0 percent absorption, it should be tested by the paraffin method (AASHTO T 275). Twenty-seven (27) samples in this experiment had absorption capacities greater than 2.0 percent, and were subsequently tested according to this method. The results of this testing are given in Table 19.

Sample Information			Gmb Values by Various Methods					
Source	NMAS	Comp. Effort	Paraffin	CL	CR	HD	KS	SSD
SS	9.5	M	2.226	2.184	2.278	2.142	2.259	2.204
SS	9.5	L	2.209	2.127	2.243	2.093	2.231	2.158
SS	9.5	L	2.182	2.137	2.234	2.100	2.227	2.159
SS	9.5	L	2.167	2.127	2.222	2.099	2.187	2.153
SS	12.5	L	2.177	2.128	2.177	2.087	2.228	2.158
SS	12.5	L	2.195	2.136	2.263	2.101	2.231	2.165
SS	12.5	L	2.198	2.147	2.260	2.117	2.219	2.179
SS	25.0	M	2.277	2.199	2.237	2.130	2.301	2.251
SS	25.0	L	2.350	2.197	2.250	2.138	2.303	2.243
SS	37.5	M	2.264	2.192	2.274	2.127	2.315	2.276
SY	25.0	L	2.303	2.214	2.148	2.153	2.306	2.266
SY	37.5	L	2.294	2.193	2.151	2.080	2.313	2.258
GR	25.0	L	2.276	2.079	2.237	2.145	2.303	2.214
GR	25.0	L	2.240	2.156	2.153	2.047	2.288	2.235
GR	25.0	L	2.206	2.118	2.177	1.979	2.265	2.214
GR	37.5	L	2.239	2.163	2.207	2.059	2.305	2.239
GR	37.5	L	2.260	2.177	2.207	2.080	2.309	2.246
GR	37.5	L	2.255	2.161	2.253	2.069	2.304	2.239
LS	25.0	M	2.342	2.260	2.282	2.181	2.375	2.295
LS	25.0	M	2.323	2.244	2.271	2.182	2.360	2.308
LS	25.0	L	2.294	2.143	2.132	2.074	2.342	2.250
LS	25.0	L	2.298	2.190	2.166	2.112	2.356	2.251
LS	25.0	L	2.288	2.174	2.268	2.098	2.353	2.258
LS	37.5	M	2.311	2.214	2.216	2.056	2.326	2.248
LS	37.5	L	2.214	2.094	2.047	1.957	2.283	2.207
LS	37.5	L	2.210	2.091	2.046	1.981	2.294	2.199
LS	37.5	L	2.249	2.131	2.181	1.949	2.298	2.207
Average Difference from Paraffin				0.091	0.047	0.167	-0.038	0.028

Table 19. Summary of Results Comparing Paraffin Gmb Values to Other Methods

A graph showing trendlines of the relationships between the paraffin and other methods is provided in Figure 17. None of the relationships were very strong, however the SSD and Kuss methods exhibited the closest relationship and were most parallel to the line of equality.

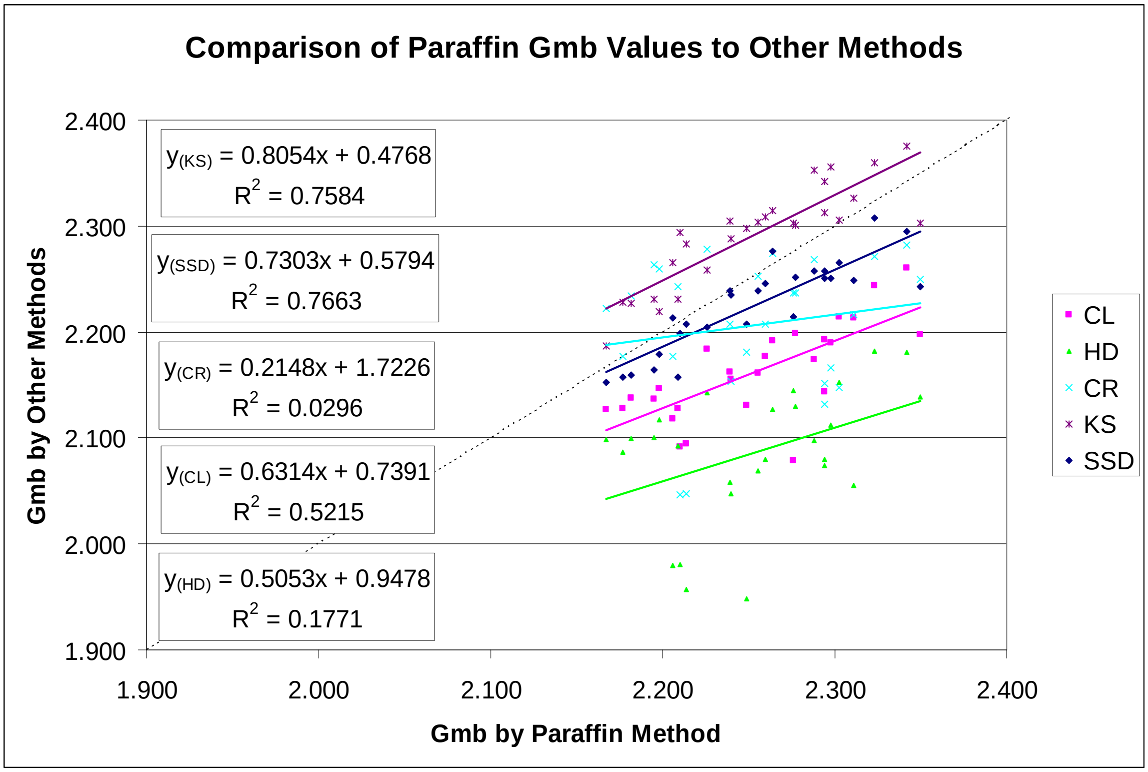

Figure 17. Relationships of Paraffin Method to Other Gmb Methods

It was expected that the densities measured by the paraffin results would be less than that of the SSD method. However, this was not true. On average, the Gmb values by the paraffin method were 0.028 larger than by the SSD method. The new results were not successful in improving the variability of the overall results, and since densities increased, it is doubtful that the accuracy was improved. The reason for this is likely the variability of the awkward and time consuming method. Also, the samples were dried after SSD testing according to the procedures outlined in AASHTO T-166. It is suspected that the sample characteristics could have changed slightly during this procedure. This is possible since most of the samples subjected to this method were those prepared with low compactive effort and may have been more fragile. Such samples may have been prone to slight consolidation during the drying process and warm temperatures.

<u>CoreReader</u>

A large amount of unexpected variability was exhibited by the CoreReader method that has not been reported in the examined literature. One possible reason for this may be sample height. The CoreReader operates in one of two modes – Lab or Field. Samples that are prepared in the laboratory should be tested in Lab mode, and cores cut from the field should be tested in Field mode. The Field mode will accept a large range of sample heights, but the Lab mode will only accept sample heights in the range of 110.0 mm – 120.0 mm. Due to the intended variations in compactive effort, some sample heights were outside this range, and had to be tested in Field mode. After discussing this problem with Troxler Laboratories, Inc., an alternative procedure was performed. The samples were re-tested in the Lab mode, based on a height of 115.0 mm. The resulting Gmb value was corrected using a proportional height calculation. This additional testing procedure did not, however, improve the results, and actually increased the variability for the CoreReader. Thus, the values generated by the corrected Lab mode procedure were not used in the remainder of the statistical analyses. It should be noted that due to this problem, the CoreReader results generated in this study may not be (and according to available literature, are not) typical.

Mathematical Modeling

Mathematical models often provide a viable option for combining the advantages of more desirable test methods with the ability to coordinate with existing specifications. By establishing a definitive relationship between traditional and alternative test methods, two benefits could be realized. First, the impacts of the alternative method on the current specifications for design and QC/QA could be predicted and assessed. This would be necessary prior to the adoption of a new test method. Secondly, if the results of various methods can be related mathematically, industry personnel could be offered the option of using alternative test methods without having to change current specifications.

The first task in this portion of the analysis was to relate the various alternative methods to the SSD method using regression tools in order to determine if a significant relationship was present. These relationships are shown in Figures 18 - 21.

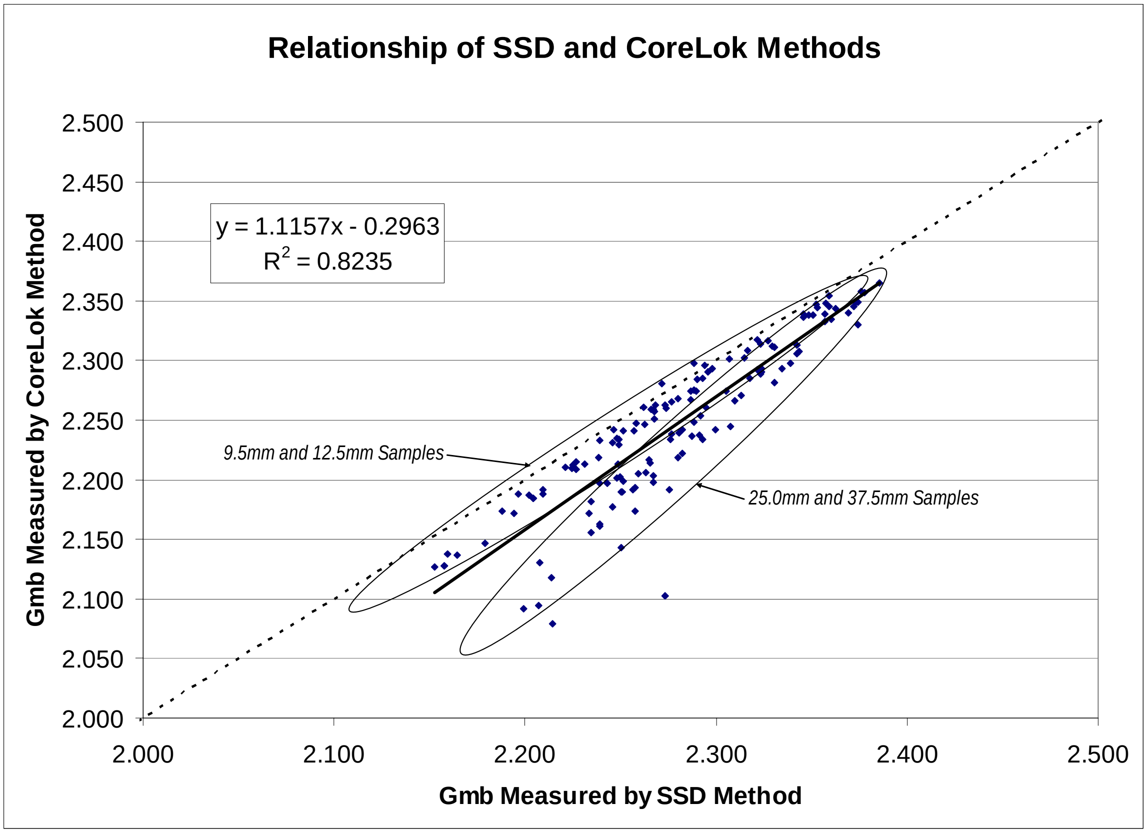

Figure 18. Relationship of SSD and CoreLok Methods

In the plot of SSD vs. CoreLok (Figure 18), a reasonable relationship was generated, and the equation was able to account for approximately 82% of the data variability as indicated by the R^2 value. This is a significant relationship, but not consistent enough for use as a prediction equation. The slope of the line was fairly parallel to the line of equality, meaning that the relationship was somewhat consistent over a range of densities. However, it was noted that the data seemed to be grouped in two separate sections. Upon further inspection, it was discovered that the upper grouping corresponded with the 9.5mm and 12.5mm samples, and the lower grouping corresponded with the 25.0mm and 37.5mm samples. This separation is consistent with the conclusions derived from the analysis of variance.

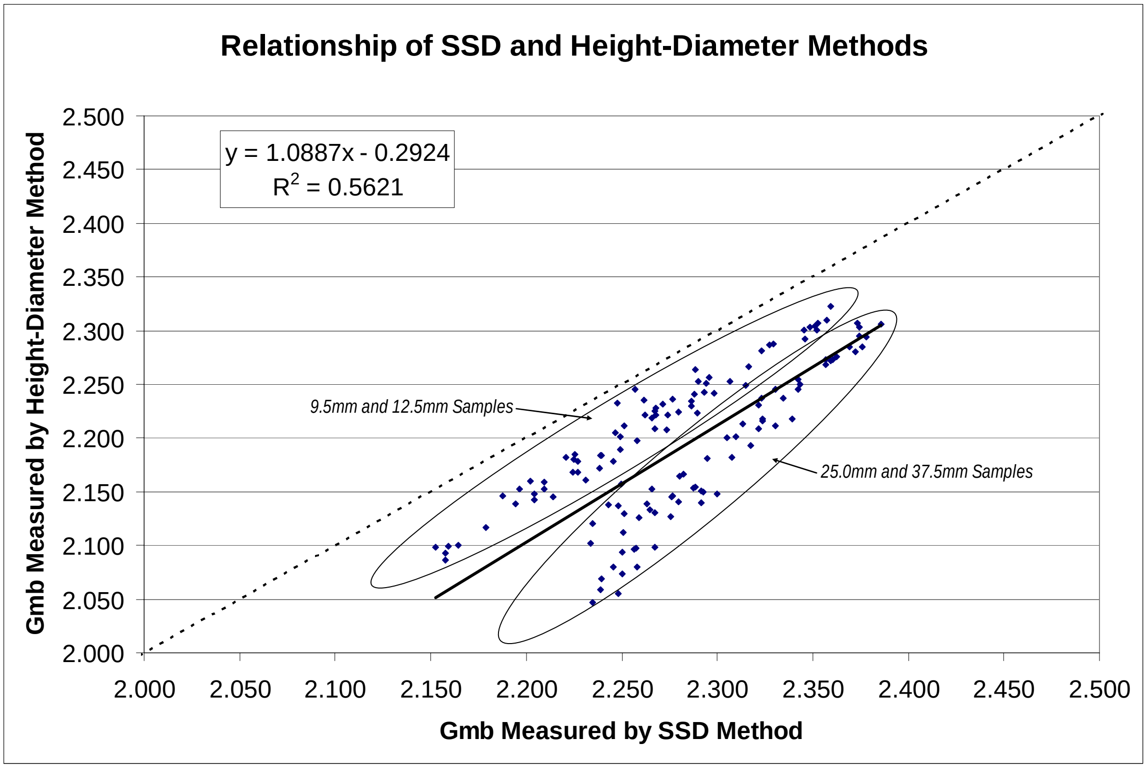

Figure 19. Relationship of SSD and Height-Diameter Methods

In the plot of SSD vs. Height-Diameter (Figure 19), a significant relationship was generated, but the equation was only able to account for approximately 56% of the data variability. The slope of the line was fairly parallel to the line of equality, meaning that the relationship was somewhat consistent over a range of densities. Again, it was noted that the data seems to be grouped in two separate sections, and as with the case of SSD vs. CoreLok, the upper grouping corresponded with the 9.5mm and 12.5mm samples, and the lower grouping corresponded with the 25.0mm and 37.5mm samples. This is also consistent with the conclusions derived from the analysis of variance.

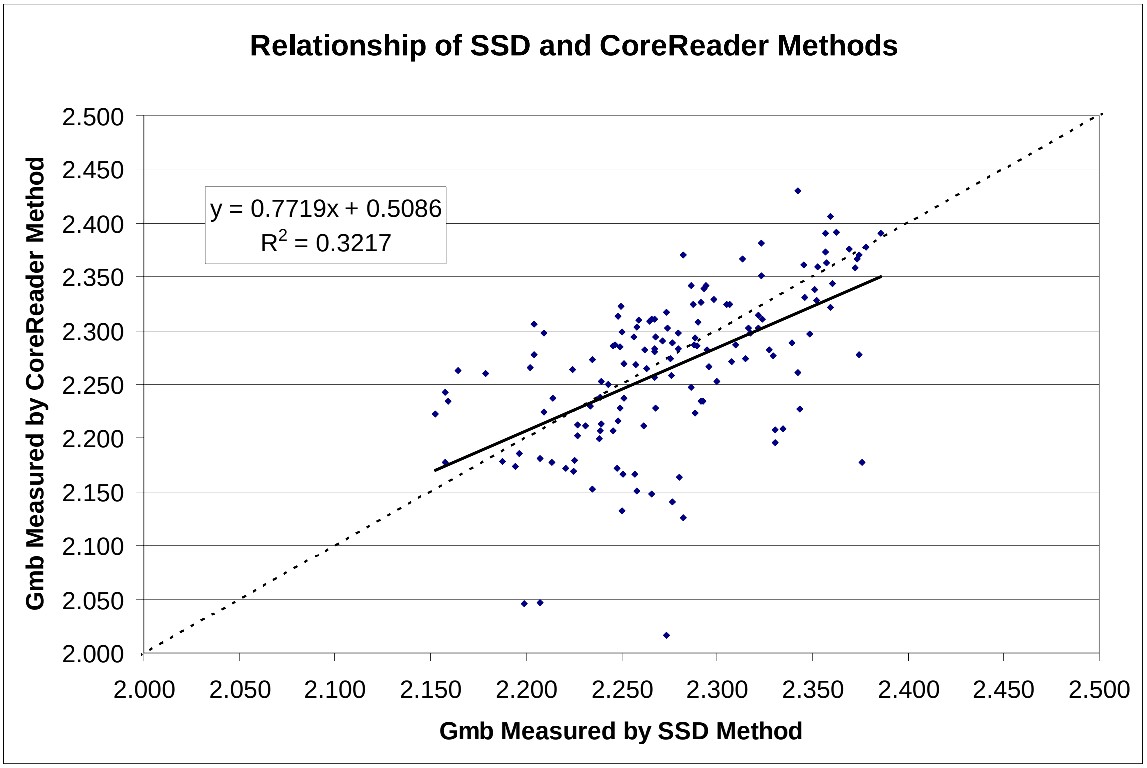

Figure 20. Relationship of SSD and CoreReader Methods

In the plot of SSD vs. CoreReader (Figure 20), the equation describing the relationship was only able to account for approximately 32% of the data variability. The data points did not display any obvious groupings, which was expected due the larger amount of data variability. One other observation was that the slope of the line that best describes the data points was not parallel to the line of equality, meaning that the relationship was not consistent as Gmb increased. For lower density samples, the CoreReader method produced larger Gmb results; for higher density samples, the SSD method produced larger Gmb results.

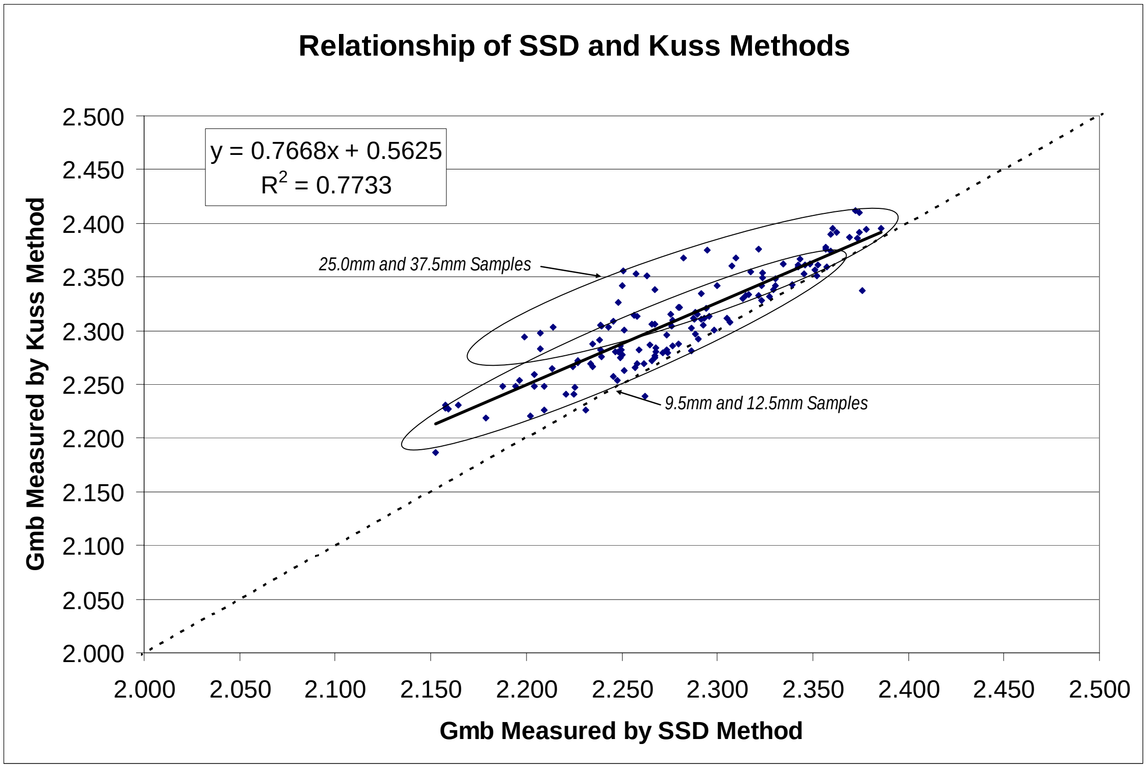

Figure 21. Relationship of SSD and Kuss Methods

In the plot of SSD vs. Kuss (Figure 21), a significant relationship was generated, and the equation was able to account for approximately 77% of the data variability. This is a significant relationship, but again, not consistent enough for use as a prediction equation. The slope of the line was not parallel to the line of equality, such that higher densities were measured by the Kuss method for samples of lower density. Although the data seemed to be somewhat grouped into sections according to aggregate size, the Kuss method was less affected by this property.

Based on these relationships, it appears that there is not yet a "magic" model to equate the various methods to the SSD method. Additional parameters are needed in order to improve the relationships. Sample characteristics were available that could be used to improve the models. However, there were very few that would be practical to include. In the study by Buchanan *(3)*, absorption capacity was determined to significantly affect the relationship between the SSD and CoreLok methods. Absorption is calculated based on weights recorded during the SSD test, so in order to use it in a predictive sense, an SSD test would have to be run in addition to the CoreLok test. This defeats the purpose of "predicting" and SSD value. In fact, the only property that could easily be incorporated is the aggregate size, and in most cases was a significant predictor value. Stepwise regression methods were used to confirm this conclusion. For the relationships of SSD vs. CoreLok and SSD vs. Height-Diameter, NMAS was the only predictor variable that significantly improved the relationships. For the SSD vs. CoreReader, NMAS and absorption value were both significant predictors. For the SSD vs. Kuss, absorption was significant, but NMAS was not. This is consistent with the lack of data grouping seen in Figure 21. For the reasons previously stated, however, absorption is not a very *practical* predictor variable.

Next, regression analyses were performed for the method comparisons based on aggregate size. In general, the 9.5mm and 12.5mm ("small") mixes were similar to each other and the 25.0mm and 37.5mm ("large") mixes were similar to each other. Thus, the samples were separated according to aggregate size – small or large. This separation did, in fact, improve the mathematical relationships. The results of this regression analysis are reported in Table 20, and shown graphically in Figures 22 – 25. Values for the intercept and coefficient of each relationship are provided, as well as the standard error for each. The adjusted R^2 value is also given, which is the R^2 value that has been adjusted (lowered) to account for the number of terms in the model. The other statistic presented is the R^2 for predictions. This term differs from the adjusted R^2 because rather than describing how well the model fits the existing data, it describes how well the model is expected to predict the responses in a *new* experiment (i.e., future data). In general, the closer the two values, the better, and the difference in the two should not be greater than 0.20 for any given model.

Method	Agg. Size	Intercept	Std. Error (Intercept)	Coefficient	Std. Error (Coefficient)	Adj. R^2	Pred. R^2
CL	Small	0.24604	0.03354	0.89660	0.01492	0.98	0.98
CL	Large	0.82756	0.05022	0.65387	0.02238	0.92	0.92
HD	Small	0.27524	0.05208	0.89817	0.02355	0.96	0.95
HD	Large	1.16448	0.05337	0.52184	0.02463	0.86	0.85
CR	Small	0.77285	0.20280	0.65551	0.08930	0.44	0.42
CR	Large	1.51514	0.12148	0.34395	0.05363	0.36	0.34
KS	Small	-0.49985	0.12485	1.20933	0.05468	0.88	0.87
KS	Large	-0.30222	0.20027	1.11212	0.08577	0.70	0.70

Table 20. Regression Analysis Results for Correlation Equations

The relationship between the SSD and CoreLok methods was very good, especially for the small aggregate sizes. The next best relationship was that for the SSD and Height-Diameter method. Again, the relationship was stronger for the small aggregate sizes. For the models relating the SSD and Kuss methods, the relationship was fairly strong for the small aggregate sizes, but not as good for the large aggregate sizes. Remember that aggregate size was not actually significant for this relationship, so the poorer correlation can be largely attributed to the greater variability of the large aggregate mixes. The CoreReader method exhibited a weak correlation, even though aggregate size was determined to be a significant factor.

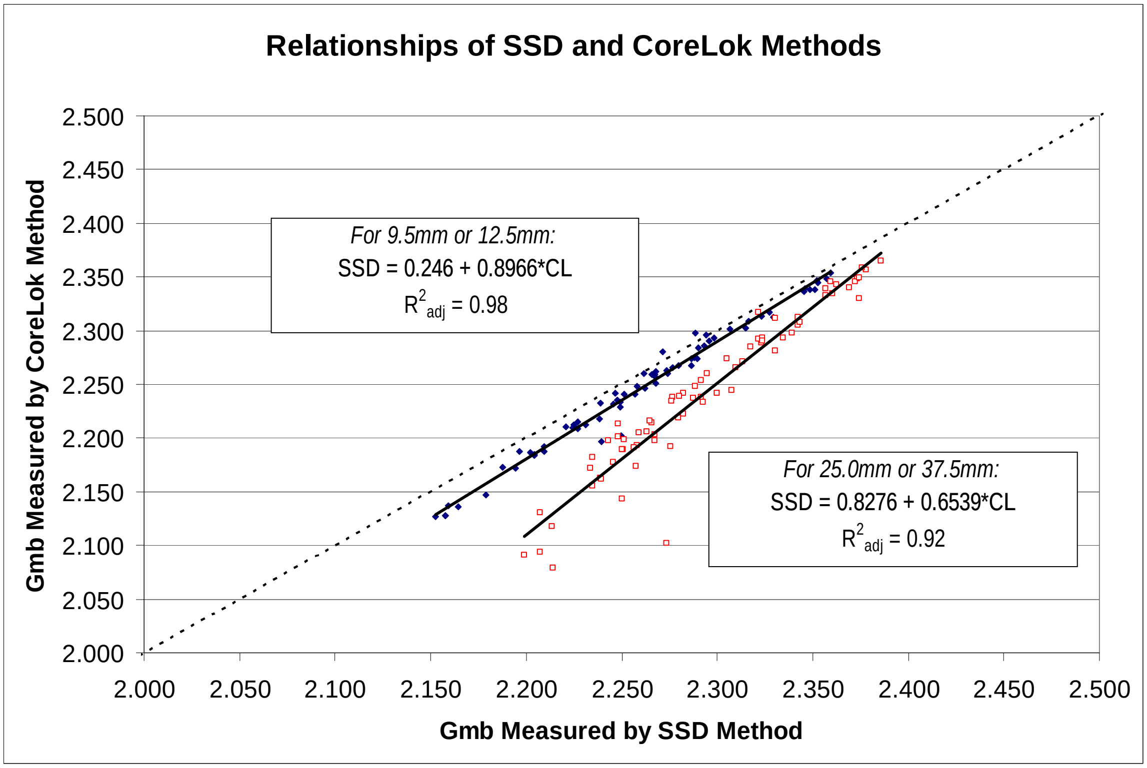

Figure 22. Relationships of SSD and CoreLok Methods by Aggregate Size

The relationships developed between the SSD and CoreLok methods were very strong when separated by aggregate size. Note that the regression line for the small aggregate mixes was much more parallel to the line of equality than that for the large aggregate mixes. This reiterates the idea that the SSD and CoreLok methods differed more for mixes of larger NMAS at lower densities. Although it is not possible to know which is more accurate, several sources have cited logical reasons for the SSD method to incorrectly estimate volumes for samples of this type. *(3, 4, 6, 7, 11, 12)*

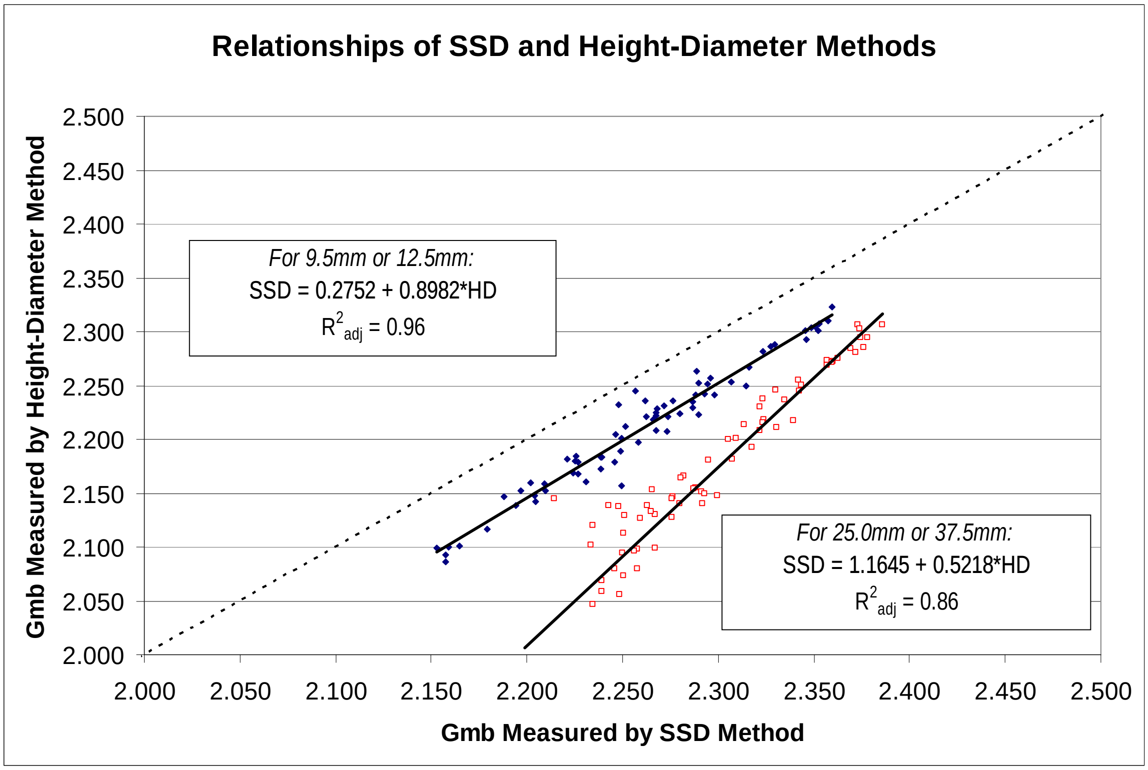

Figure 23. Relationships of SSD and Height-Diameter Methods by Aggregate Size

In the models relating the SSD and Height-Diameter methods, the regression line was quite parallel to the line of equality for the small mixes and non-parallel for the large mixes. Although the distances of these lines from the line of equality was greater than that for the CoreLok comparison, similar trends emerged. As the NMAS increased and densities decreased, the measures of Gmb by the SSD and Height-Diameter methods became “more different”.

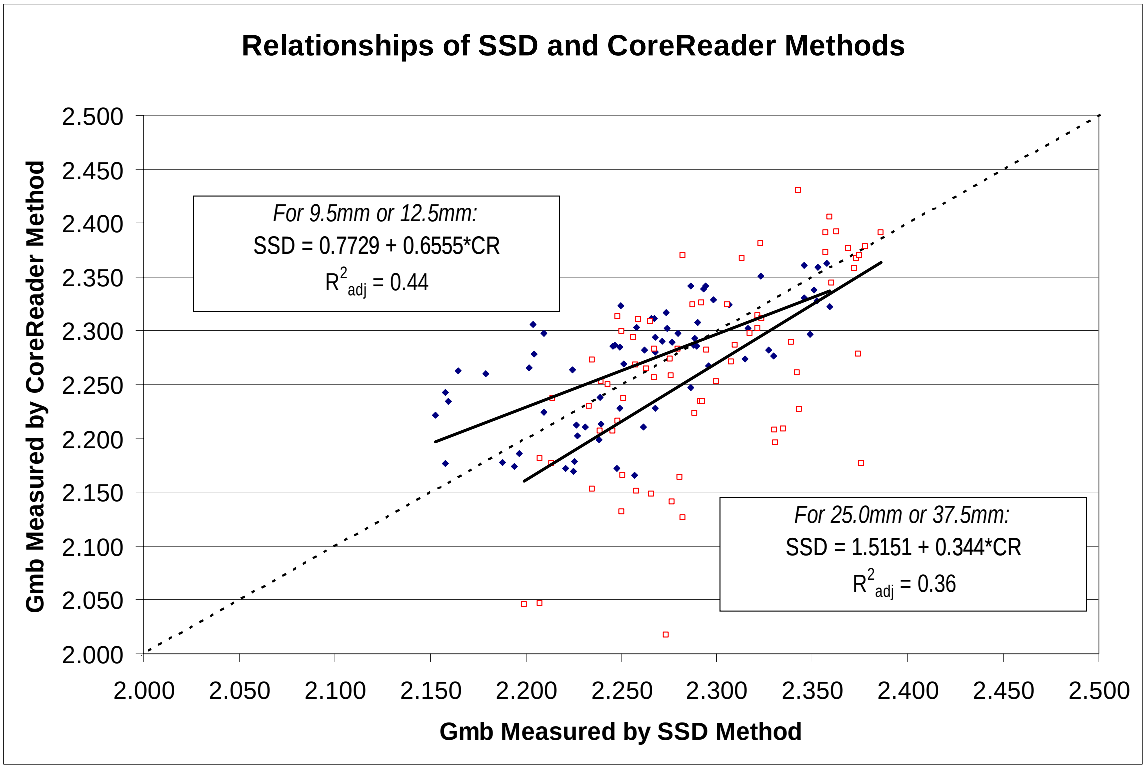

Figure 24. Relationships of SSD and CoreReader Methods by Aggregate Size

In the CoreReader relationships, there was little to conclude due to the weak data relationships. The variability within the data is the primary cause for this problem. It is interesting to note, however, that the regression line for the larger aggregate mixes was more parallel to the line of equality than for the smaller mixes.

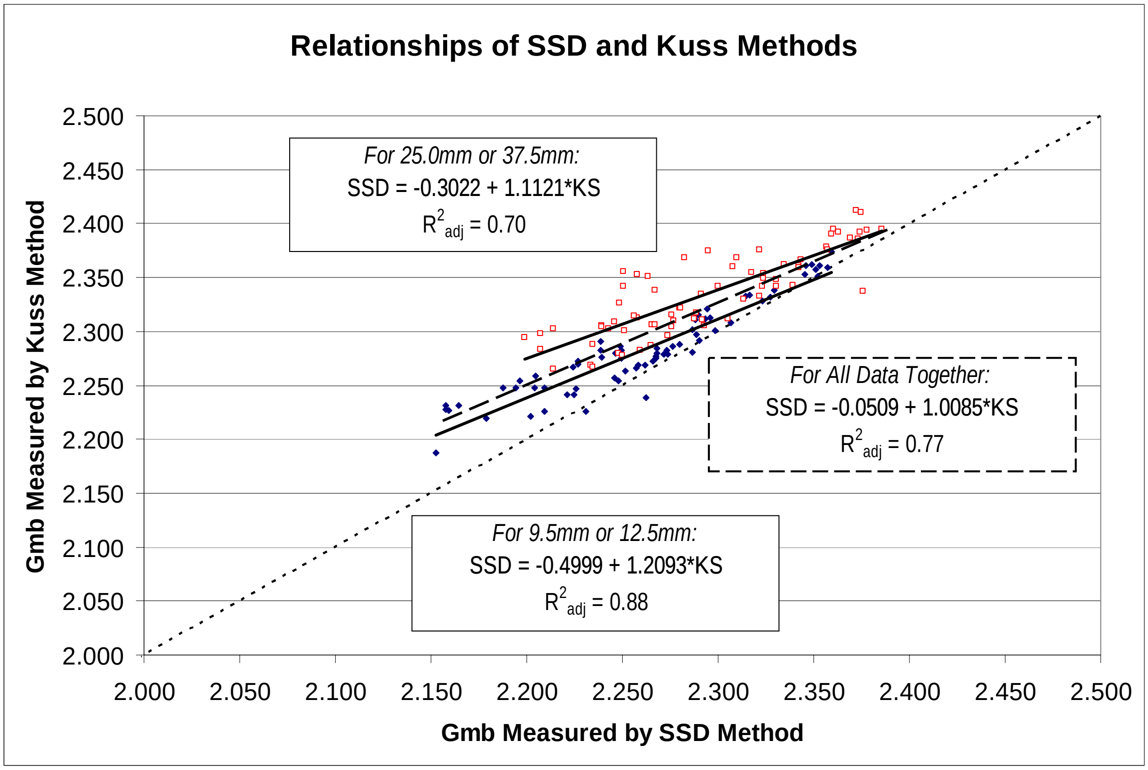

Figure 25. Relationships of SSD and Kuss Methods by Aggregate Size

For the SSD and Kuss method relationships, a fair correlation was obtained. The aggregate size term was not actually significant, which is evidenced here by the fact that the slopes of the two lines for separate aggregate sizes were very similar. The dashed line represents the model for all data (not separated by aggregate size).

Absorption was a significant factor in the relationship between the SSD and Kuss methods. When this term was added to the model, the correlation was greatly improved, having adjusted R^2 and prediction R^2 values of 0.94. The resulting relationship is given in Equation 6.

$$SSD = 0.3047 + 0.869*KS - 0.0263*ABSORPTION \qquad \textit{Equation 6}$$

Even though this is an excellent relationship, the SSD method must be performed first in order to calculate absorption. As previously discussed, this negates the purpose of substituting an alternative test method. However, the Kuss method involves placing the specimen in water, so if the procedure could be altered in such a way to provide a value for absorption, then this relationship could be more meaningful.

Validation

Once relationships have been developed, they should be validated using separate, independent sample sets. Two sample sets were used to validate the equations. The first included 120 samples from five 12.5mm surface mixes from various sources in Arkansas. The plant mix for each job was obtained such that six samples were compacted from each of four separate sublots during that job, comprising a total of 24 samples per job. The mix was then brought back to the laboratory and compacted in the gyratory compactor. The second sample set included 39 12.5mm samples and 38 25.0mm samples (one 25.0mm sample was damaged before testing was complete). All 77 samples were batched, mixed, and compacted in the laboratory to various levels of compaction.

In this analysis, the Gmb for each sample was tested according to the SSD, CoreLok, and Height-Diameter methods. The CoreLok and Height-Diameter results were used with the appropriate mathematical models to calculate a predicted SSD value. The predicted and actual SSD densities were then compared.

According to AASHTO T-166, “two results should not differ by more than 0.02”. *(1)* This is the traditional established value for the acceptable range of two results (d2s) by the same operator for the SSD method. Therefore, the predicted and actual SSD Gmb values were said to be “different” if the absolute value of the difference was greater than 0.02. The results are summarized in Table 21.

		d2s = 0.02			
		12.5mm Plant Mix	12.5mm Lab Mix	25.0mm Lab Mix	All Samples
Prediction based on CoreLok	# of Diff. Results	12	1	6	19
	# of Comparisons	120	39	38	197
	% Different	10.0	2.6	15.8	9.6
Prediction based on Ht.-Dia.	# of Diff. Results	48	11	13	72
	# of Comparisons	120	39	38	197
	% Different	40.0	28.2	34.2	36.5

Table 21. Summary of Results Predicted vs. Actual SSD (d2s = 0.02)

With the exception of the 12.5mm laboratory mix, an appreciable number of differences were detected. Thus, use of the models is questionable. Some of the differences are likely due to the variability of the alternative test methods, which have been shown to be greater than that of the SSD method. *(11, 12)*

In the round robin study, NCAT indicated that the within lab d2s value for the SSD method should actually be 0.052, which is much larger than the value of 0.02 published in the AASHTO standard. *(1)* The range of results from several round robin studies was 0.026 to 0.052. If this is the case, then a mid-range d2s value of 0.04 may be more appropriate. Therefore, the validation results were re-evaluated on the basis of a d2s value of 0.04, and much better agreement was obtained. The summary of results is given in Table 22.

		d2s = 0.04			
		12.5mm Plant Mix	12.5mm Lab Mix	25.0mm Lab Mix	All Samples
Prediction based on CoreLok	# of Diff. Results	1	1	0	2
	# of Comparisons	120	39	38	197
	% Different	0.8	2.6	0.0	1.0
Prediction based on Ht.-Dia.	# of Diff. Results	13	3	5	21
	# of Comparisons	120	39	38	197
	% Different	10.8	7.7	13.2	10.6

Table 22. Summary of Results Predicted vs. Actual SSD (d2s = 0.04)

Variability Based on Aggregate Size

Because most of the mathematical relationships are significantly affected by aggregate size, the variability data was separated and re-evaluated to consider variability in terms of small and large aggregate mixes. The variability summaries for small and large aggregate mixes are given in Tables 23 and 24, respectively.

			Standard Deviation and Coefficient of Variation (%)									
			CoreLok		CoreReader		Height-Diameter		Kuss		SSD	
Source	NMAS	Comp. Effort	Std. Dev.	COV%	Std. Dev.	COV %	Std. Dev.	COV%	Std. Dev.	COV%	Std. Dev.	COV%
SS	9.5	High	0.0066	0.291	0.0046	0.200	0.0105	0.474	0.0030	0.132	0.0050	0.220
SS	9.5	Medium	0.0039	0.178	0.0284	1.257	0.0097	0.449	0.0206	0.924	0.0037	0.170
SS	9.5	Low	0.0058	0.274	0.0105	0.472	0.0038	0.181	0.0243	1.098	0.0034	0.158
SS	12.5	High	0.0343	1.532	0.0108	0.467	0.0338	1.542	0.0017	0.076	0.0138	0.611
SS	12.5	Medium	0.0262	1.189	0.0101	0.438	0.0167	0.775	0.0052	0.231	0.0227	1.022
SS	12.5	Low	0.0094	0.439	0.0488	2.185	0.0152	0.724	0.0062	0.281	0.0109	0.503
SY	9.5	High	0.0092	0.402	0.0155	0.672	0.0062	0.275	0.0082	0.356	0.0100	0.437
SY	9.5	Medium	0.0139	0.608	0.0228	0.991	0.0052	0.230	0.0112	0.491	0.0142	0.621
SY	9.5	Low	0.0084	0.372	0.0070	0.307	0.0068	0.305	0.0192	0.850	0.0083	0.365
SY	12.5	High	0.0068	0.302	0.0667	2.919	0.0065	0.290	0.0096	0.422	0.0128	0.562
SY	12.5	Medium	0.0143	0.637	0.0342	1.562	0.0088	0.396	0.0130	0.574	0.0101	0.445
SY	12.5	Low	0.0010	0.045	0.0051	0.236	0.0022	0.103	0.0035	0.154	0.0025	0.112
GR	9.5	High	0.0024	0.103	0.0414	1.795	0.0034	0.148	0.0050	0.216	0.0032	0.136
GR	9.5	Medium	0.0110	0.493	0.0047	0.206	0.0111	0.500	0.0045	0.198	0.0115	0.512
GR	9.5	Low	0.0089	0.407	0.0061	0.280	0.0069	0.323	0.0035	0.154	0.0045	0.206
GR	12.5	High	0.0059	0.253	0.0020	0.085	0.0047	0.203	0.0042	0.177	0.0060	0.255
GR	12.5	Medium	0.0106	0.465	0.0315	1.356	0.0144	0.641	0.0046	0.198	0.0024	0.106
GR	12.5	Low	0.0182	0.822	0.0255	1.139	0.0088	0.403	0.0075	0.332	0.0085	0.380
LS	9.5	High	0.0091	0.390	0.0207	0.891	0.0109	0.474	0.0087	0.370	0.0056	0.239
LS	9.5	Medium	0.0119	0.523	0.0200	0.882	0.0135	0.600	0.0059	0.254	0.0049	0.214
LS	9.5	Low	0.0128	0.577	0.0131	0.592	0.0106	0.486	0.0087	0.383	0.0129	0.576
LS	12.5	High	0.0051	0.219	0.0021	0.091	0.0062	0.270	0.0071	0.300	0.0045	0.192
LS	12.5	Medium	0.0043	0.188	0.0198	0.865	0.0124	0.547	0.0007	0.030	0.0011	0.049
LS	12.5	Low	0.0039	0.176	0.0085	0.385	0.0080	0.371	0.0460	2.035	0.0052	0.231
AVERAGE VALUES			0.0102	0.453	0.0192	0.845	0.0098	0.446	0.0097	0.426	0.0078	0.347

Table 23. Standard Deviation and Coefficient of Variation (%) for Small Aggregate Mixes

			Standard Deviation and Coefficient of Variation (%)									
			CoreLok		CoreReader		Height-Diameter		Kuss		SSD	
Source	NMAS	Comp. Effort	Std. Dev.	COV%	Std. Dev.	COV %	Std. Dev.	COV%	Std. Dev.	COV%	Std. Dev.	COV%
SS	25.0	High	0.0096	0.418	0.0361	1.545	0.0198	0.892	0.0154	0.661	0.0099	0.429
SS	25.0	Medium	0.0034	0.155	0.0384	1.692	0.0050	0.235	0.0365	1.578	0.0080	0.354
SS	25.0	Low	0.0127	0.582	0.0215	0.956	0.0182	0.856	0.0202	0.887	0.0051	0.228
SS	37.5	High	0.0169	0.750	0.0243	1.038	0.0390	1.798	0.0107	0.461	0.0139	0.603
SS	37.5	Medium	0.0134	0.609	0.0052	0.228	0.0068	0.317	0.0080	0.347	0.0065	0.284
SS	37.5	Low	0.0135	0.612	0.0061	0.264	0.0207	0.976	0.0045	0.198	0.0075	0.332
SY	25.0	High	0.0053	0.226	0.1130	4.898	0.0109	0.475	0.0309	1.301	0.0024	0.101
SY	25.0	Medium	0.0159	0.691	0.0346	1.557	0.0232	1.035	0.0097	0.413	0.0067	0.288
SY	25.0	Low	0.0121	0.543	0.1302	5.868	0.0100	0.464	0.0347	1.490	0.0085	0.372
SY	37.5	High	0.0178	0.760	0.0614	2.613	0.0116	0.506	0.0040	0.169	0.0085	0.359
SY	37.5	Medium	0.0085	0.369	0.1930	8.288	0.0160	0.718	0.0093	0.394	0.0098	0.421
SY	37.5	Low	0.0697	3.198	0.0814	3.856	0.0961	4.636	0.0132	0.571	0.0116	0.510
GR	25.0	High	0.0175	0.758	0.0709	3.000	0.0253	1.128	0.0175	0.742	0.0095	0.404
GR	25.0	Medium	0.0069	0.306	0.0685	3.110	0.0069	0.321	0.0092	0.398	0.0062	0.272
GR	25.0	Low	0.0385	1.820	0.0433	1.977	0.0832	4.046	0.0191	0.838	0.0120	0.540
GR	37.5	High	0.0031	0.134	0.0084	0.350	0.0020	0.090	0.0087	0.365	0.0028	0.121
GR	37.5	Medium	0.0041	0.183	0.0110	0.490	0.0015	0.070	0.0197	0.845	0.0045	0.197
GR	37.5	Low	0.0090	0.413	0.0266	1.195	0.0108	0.522	0.0026	0.115	0.0039	0.173
LS	25.0	High	0.0073	0.311	0.0130	0.552	0.0113	0.495	0.0093	0.386	0.0077	0.325
LS	25.0	Medium	0.0113	0.500	0.0082	0.359	0.0115	0.526	0.0075	0.317	0.0081	0.350
LS	25.0	Low	0.0236	1.090	0.0708	3.234	0.0197	0.939	0.0074	0.314	0.0041	0.182
LS	37.5	High	0.0172	0.747	0.0067	0.289	0.0117	0.531	0.0124	0.526	0.0031	0.132
LS	37.5	Medium	0.0113	0.513	0.0390	1.729	0.0243	1.164	0.0120	0.516	0.0094	0.417
LS	37.5	Low	0.0219	1.038	0.0777	3.713	0.0168	0.858	0.0078	0.339	0.0049	0.221
AVERAGE VALUES			0.0154	0.697	0.0495	2.200	0.0209	0.983	0.0138	0.590	0.0073	0.317

Table 24. Standard Deviation and Coefficient of Variation (%) for Large Aggregate Mixes

It is especially interesting to note that the SSD Method coefficient of variation was slightly smaller for the large aggregate mixes than for the small aggregate mixes. For all other methods, the variability associated with the large aggregate mixes was larger than that for the small aggregate mixes. The greatest difference was evident for the CoreReader method.

Mix Design

One way to use these relationships is to evaluate mix designs. Using correlated values for alternative test methods in a mix design can aid in assessing the impact that the new method could have on design specifications.

For example, assume that the following mixture was designed using SSD Gmb values.

- NMAS = 25.0mm
- Design Asphalt Content = 4.2%
- Design Air Content = 4.5%
- Design VMA = 14.1%
- Design VFA = 69.0%
- Dust Proportion = 0.8
- % Passing #200 = 3.5%
- Gmm = 2.392
- Gmb @ Ndes = 2.284

The relationship of SSD and CoreLok for large aggregate mixes can be used to estimate what the Gmb *would have been* if measured by the CoreLok method. In this case, that value would be 2.227. The calculated air voids, then, would become 6.9 percent. This change in Gmb has a huge practical significance in terms of air voids – almost 2.5 percent. So if it is assumed that the CoreLok provides a more "accurate" measure of bulk density than the SSD method, then what seemed like a mixture designed at 4.5 percent air voids was actually designed at 6.9 percent air voids. The change in Gmb also has a very significant effect on the VMA percentage. A selection of design graphs for the SSD case and CoreLok case are presented in Figures 26 and 27, respectively.

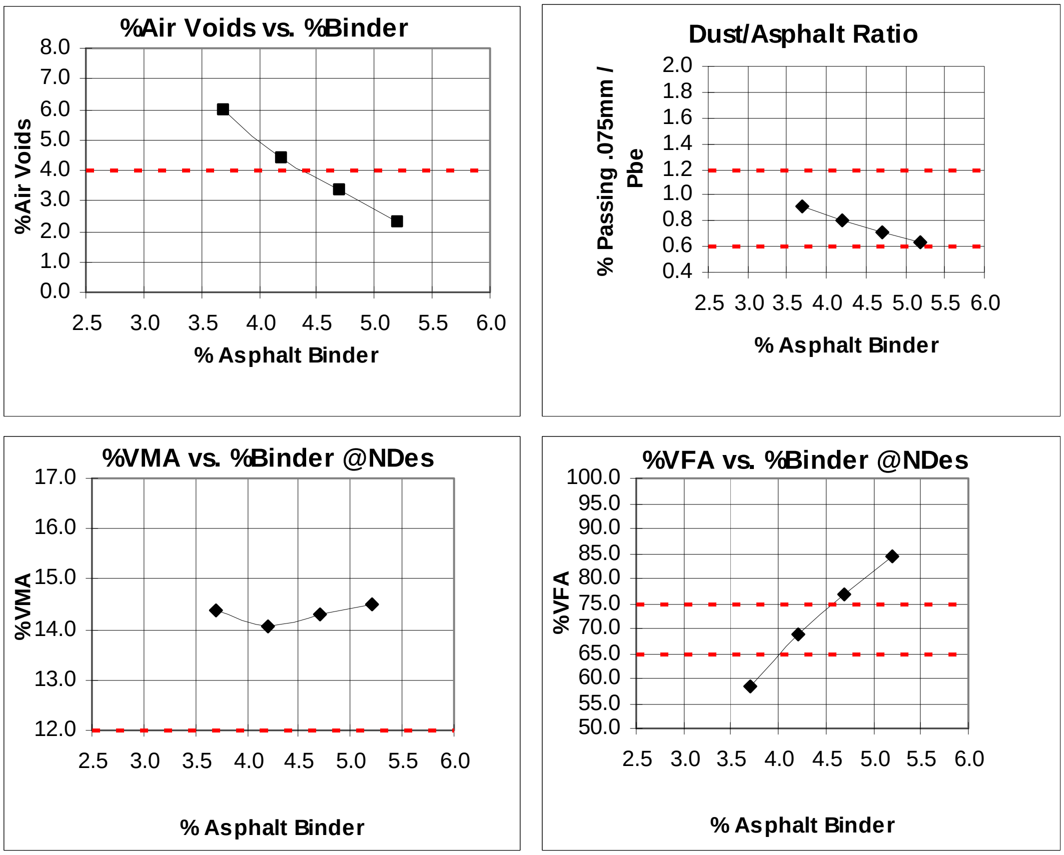

Figure 26. Design Graphs Using Measured SSD Gmb Values

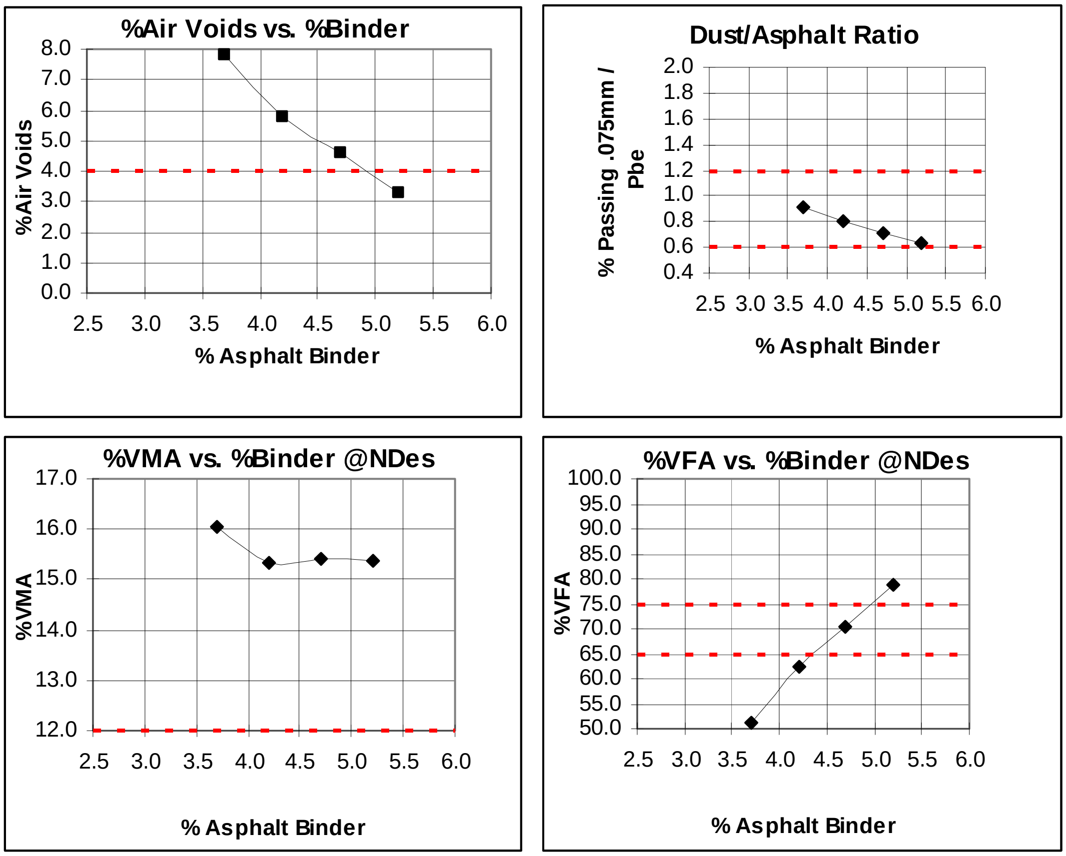

Figure 27. Design Graphs Using SSD Values Estimated from CoreLok Gmb Values

Based on a design air void content of 4.5 percent, the design binder content changed to 4.7 percent – an increase of 0.5 percent. The VMA and VFA increased to 15.4 percent and 70.4 percent, respectively, and the dust proportion decreased to 0.7. The problem with this design, however, is that the new design binder content is now on the "wet" side of the VMA curve. Thus, the aggregate blends would have to be adjusted to close the VMA without using excess asphalt cement. This might be accomplished by adding fines. As fines are added, the mix becomes finer, more of the large air void spaces are removed, and the interconnectivity of the void spaces is reduced. This type of mix design might resemble those designed many years ago.

It is important to consider the fact that the slope of the relationship between the SSD and CoreLok methods was not parallel to the line of equality. As densities increased, the line describing the relationship approached the line of equality. Therefore, as the density of the materials used in the mix design increased, the discrepancy would decrease. If the correlation line were extrapolated so that it crosses the line of equality, the difference in SSD and CoreLok Gmb values would be minimized between densities of 2.400 and 2.450, which is greater than that of the samples tested in this research. Therefore, mix designs containing higher density materials might be less affected by changing to the CoreLok method than those containing lower density materials. The range of densities tested in this study is typical of Arkansas aggregates and mixtures, and may not reflect the typical densities of materials in other states.

Field Cores

In order to reap the benefits of an alternative test method, it must be applicable to both laboratory-compacted and field-compacted samples. Gyratory-compacted samples contain a density-gradient, having a higher density in the interior portion of the sample and lower densities on the exterior portions. *(11)* This is generally attributed to the difference in confining conditions of laboratory and field compaction methods. The lower densities near the exterior of the laboratory-compacted samples creates more surface irregularities, which are often blamed for difficulties in determining accurate volumes in the SSD and Height-Diameter methods. Field-compacted samples are more consistent, so it is easier to obtain and "accurate" volume according to the SSD method.

Sixteen pavement cores were cut from two jobs, and measured according to the five Gmb methods. Ten cores were cut from a 12.5mm Limestone mix, and six cores were cut from a 25.0mm Gravel mix. A summary of the data is presented in Table 25.

According to these summary values, the Kuss method provided the largest mean densities, followed by the SSD, CoreLok, CoreReader, and Height Diameter methods. These trends were consistent with that seen in the analysis involving only laboratory samples. In terms of variability, the Kuss and SSD methods were least variable, while the CoreReader and Height-Diameter methods were most variable. The Height-Diameter method might be expected to be less variable since the density of a field core is more consistent than a laboratory specimen, thus reducing the amount of surface irregularities present. However, other types of surface irregularities could be present in the form of ridges or non-uniformities caused by the drilling and removal of the core from the mat. Therefore, it is not surprising that this method exhibited a relatively high variability. The overall levels of variability were larger for the field cores than for the laboratory-compacted samples. However, a comparison of laboratory and field variability is not appropriate in this case because 1) the number of samples tested was very different, and 2) the field samples were obtained from different locations along the job, having an additional unknown variability component.

		Summary Statistics		
	Method	Average	Standard Dev.	COV (%)
Mix 1 12.5mm Limestone	CL	2.246	0.0413	1.838
	CR	2.196	0.078	3.561
	HD	2.185	0.0556	2.546
	KS	2.295	0.0291	1.268
	SSD	2.263	0.0362	1.600
Mix 2 25.0mm Gravel	CL	2.359	0.0202	0.857
	CR	2.344	0.0299	1.277
	HD	2.289	0.0332	1.449
	KS	2.367	0.0203	0.857
	SSD	2.358	0.0196	0.831
TOTAL	CL	2.288	0.0308	1.348
	CR	2.252	0.0541	2.419
	HD	2.224	0.0444	1.997
	KS	2.322	0.0247	1.062
	SSD	2.299	0.0279	1.216

Table 25. Field Sample Data Summary

The field data was tested by ANOVA using a complete randomized block design with a single factor. The analysis revealed that the block was significant, and there was also a significant difference between methods. The results of the ANOVA and means test summary are given in Tables 26 and 27.

Factor	df	F-value	P-value
Method	4	12.51	<0.0001
Source	1	108.53	<0.0001
Error	74	MSE = 0.001962	

Table 26. Field Sample ANOVA Summary

	Duncan's Multiple Range Test Results				
Method	KS	SSD	CL	CR	HD
Mean	2.322	2.299	2.288	2.252	2.224
Rank	**A**	**AB**	**B**	**C**	**C**

Table 27. Field Sample Means Test Summary

Next, the prediction equations for the CoreLok and Height-Diameter methods were used to assess their validity for the field cores. After all, if the prediction equations are to be used, they should be deemed appropriate for all sample types (i.e., lab or field). To do this, CoreLok and Height-Diameter Gmb values were used to predict SSD values and compare them to the actual SSD values. Based on the values for an acceptable range of two results (within-lab d2s) of 0.02 and 0.04, good results were obtained for the CoreLok relationships, but were only fair for the Height-Diameter relationships. Table 28 summarizes the results.

		d2s = 0.02		*d2s = 0.04*	
		12.5mm Field Cores	25.0mm Field Cores	12.5mm Field Cores	25.0mm Field Cores
Prediction based on CoreLok	# of Diff. Results	0	1	0	0
	# of Comparisons	10	6	10	6
	% Different	0.0	16.7	0.0	0.0
Prediction based on Ht.-Dia.	# of Diff. Results	6	1	2	0
	# of Comparisons	10	6	10	6
	% Different	60.0	16.7	20.0	0.0

Table 28. Percent of Different Results Using Predictive Equations for Field Cores

Field Density

Another topic that should be addressed is the effect of various Gmb test methods on field density measurements made by the nuclear gauge. In order to use the nuclear gauge to measure mat density, it must first be calibrated for the mix. In Arkansas, a job correction factor must be developed based on the average difference in nuclear gauge readings and core densities at five locations, averaging four gauge readings per core. This correction factor is then applied to all nuclear gauge readings. Obviously, the accuracy of the nuclear gauge is dependent upon the accuracy of the measured core densities. Therefore, the method with the greatest accuracy and the lowest variability is the best choice.

The SSD method is suspected to overestimate density for open and coarse-graded samples. The specification for field compaction in Arkansas requires 92 to 96 percent density, and typical actual densities hover close to the minimum value. This means that most field cores can be expected to contain seven to eight percent air voids, which is exactly the type of sample that is most difficult to measure by the SSD method.

Consider the following example. A field core is tested by the SSD method and has a density of 2.279. The Gmm value for the 25.0mm mix is 2.464, and therefore the calculated field density is 92.5 percent. If the CoreLok method is assumed to be the most accurate, then the correlation equation can be used to determine that the corresponding CoreLok density would be 2.220, which translates to only 90.1 percent density. This difference (2.4 percent) represents a huge impact to the contractor. Of course, if the CoreLok were used to measure field density, then it should also be used to measure Gmb during the design of the mix. Thus the design of the mix would be different, and the contractor may not really have to increase compaction by three percent.

Next, consider this mix from a design standpoint. Assume that the mix was designed at an air void content of 4.5 percent, having a Gmb (by SSD method) of 2.353. This Gmb correlates to a CoreLok Gmb value of 2.333, which causes the calculated air void content to become 5.3 percent. The difference in air voids is 0.8 percent during design, and 2.4 percent during construction. A discrepancy between design and construction of 1.6 percent air voids is present because the difference in the CoreLok and SSD methods is greater at lower densities, especially for the large aggregate mixes. And, because mixes are placed at lower densities than they are designed, the impact of adopting the CoreLok method for design could be very significant.

DISCUSSION

Because SSD Gmb values are currently used in all phases of asphalt mix design and production, and specification requirements are based on these values, it is sensible to continue using the SSD method until another method provides a substantial advantage such as greater accuracy, greater precision, or a significant savings of money and/or time. In terms of accuracy, it is difficult to suggest that one method has a greater accuracy than another because there is no "true" method for determining Gmb. Hence, there is no absolute measure and all comparisons are relative. The CoreLok method produced densities lower than the SSD method (which is believed to overestimate densities) and higher than the Height-Diameter method (which is believed to underestimate densities). Thus, the CoreLok method may be more accurate. It has been demonstrated that other methods do not provide results equivalent to that of SSD, implying a potentially significant impact to specifications for asphalt design and production.

In terms of precision, definitive comparisons can and have been made. *(3, 10, 11, 12)* Although this study determined that the SSD method is the least variable method, existing literature indicates that other methods may possess the ability to measure Gmb with similar or lower variability than the SSD method. Most of these studies have been limited to smaller aggregate sizes, which (in this study) tend to be more consistent than the larger ones. This research scope included a larger range of aggregate sizes than those found in the literature. This study did not, however, analyze 19.0mm mixtures. For agencies using such mixes, further testing should be performed to assess the applicability of these conclusions.

Another point to consider is the difference in repeatability and reproducibility. The method variability values found in this research are associated with single-operator variability. The reproducibility value is also important to consider since multiple operators perform QC and QA testing on typical jobs. The AMRL study discovered that although the SSD exhibited the lowest overall multi-operator variability, a greater portion of that quantity was due to the multiple operator for the SSD method than for the CoreLok method. *(12)* The CoreLok method, then, has the advantage that multiple operators can perform tests almost as precisely as a single operator. In other words, the CoreLok method is relatively operator-independent.

CONCLUSIONS

The following conclusions and recommendations are made with regard to the results of this study.

Gmm

- AASHTO T-209 and the CoreLok methods for measuring Gmm provided similar results.
- The Kuss method yielded statistically higher densities.
- AASHTO T-209 had the lowest variability, followed by the CoreLok and Kuss methods, respectively.

It is recommended that the CoreLok and Kuss methods be further evaluated, focusing on ways to reduce variability. AASHTO T-209 should continue to be used in current methods and specifications for the design and quality control / quality assurance of HMA mixtures.

Gmb

- The various methods for measuring Gmb did not yield similar results.
- The Kuss method generated the highest densities, followed by the SSD method. The Height-Diameter method generated the lowest densities, and the CoreLok method produced mid-range values.
- Gmb values were significantly affected by test method, NMAS, and compactive effort, as well as combinations of these factors.
- The Height-Diameter method was most sensitive to NMAS.
- The CoreLok and Kuss methods were least sensitive to NMAS.
- The Kuss method was least sensitive to compactive effort.
- Small aggregate sizes were less affected by compactive effort.

- When analyzed by mix, the CoreLok and SSD methods were similar in 20 of 24 cases involving small aggregate sizes (9.5mm and 12.5mm), and in only 7 of 24 cases involving large aggregate sizes (25.0mm and 37.5mm).
- The difference in the Height-Diameter and SSD methods was most sensitive to NMAS, whereas the difference in the Kuss and SSD methods was least sensitive to NMAS.
- The SSD was the least variable method tested, followed by the Kuss method, the CoreLok method, the Height-Diameter method, and the CoreReader method.
- Gmb values were more variable as NMAS increased and compactive effort decreased.
- Significant mathematical relationships were developed to relate the various test methods to the SSD method.
- The best and most practical correlations were developed for the CoreLok method.
- Stronger mathematical correlations were developed for the smaller values of NMAS.
- The mathematical correlations were validated for laboratory-compacted samples, plant-produced/laboratory-compacted samples, and field cores.
- Mix design specifications can be significantly impacted by the use of alternative test methods for measuring Gmb.
- Using the CoreLok for mixture design could cause design asphalt contents to increase, and/or gradations to shift toward finer blends.
- Using the CoreLok for field core measurements could force the contractor to meet higher density requirements.

Based on the topics evaluated in this research, changing current HMA specifications to accommodate new test methods for Gmb and Gmm is not yet warranted. While alternative test methods do possess some advantages, the elimination of traditional test methods is discouraged.

REFERENCES

1. *Standard Specifications for Transportation Materials and Methods of Sampling and Testing*, American Association of State Highway and Transportation Officials, 25th Edition, Washington, D.C., 2005.

2. *Standard Specification for Highway Construction*, Arkansas State Highway and Transportation Department, Little Rock, Arkansas, 2003.

3. Buchanan, M. Shane, "An Evaluation of Selected Methods for Measuring the Bulk Specific Gravity of Compacted Hot Mix Asphalt (HMA) Mixes," *Journal of the Association of Asphalt Paving Technologists*, Vol. 69, March 2000.

4. Crouch, L.K., Copeland, A.R., Walker, C.T., Maxwell, R.A., Duncan, G.M., Goodwin, W.A., Badoe, D.A., and Weimer, H.W., "Determining Air Void Content of Compacted HMA Mixtures, *Transportation Research Record*, Transportation Research Board, Washington, D.C., 2002.

5. "Superpave Level 1 Mix Design", Superpave Series No. 2 (SP-2), Asphalt Institute, Lexington, Kentucky, 1995.

6. "Measuring Bulk Specific Gravity of compacted Specimens Using the Troxler Model 3663 CoreReader", Application Brief, Troxler Electronic Laboratories, Inc., Research Triangle Park, NC, March, 2001.

7. Hall, K.D., Griffith, F.T., and Williams, S.G., "Examination of Gamma-Ray Method for Measuring Bulk specific Gravity of Hot-Mix Asphalt Concrete", *Transportation Research Record #1832*, Transportation Research Board, Washington, D.C., 2003.

8. Rice, J.M., "New Test Method for Direct Measurement of Maximum Density of Bituminous Mixtures", *Crushed Stone Journal*, September 1953.

9. Bukowski, J.R., (editor), Meeting minutes – TRB Superpave Mixture and Aggregate Expert Task Group, March 2000, Transportation Research Board, Washington, D.C.

10. Hall, K.D., Griffith, F.T., and Williams, S.G., "An Examination of Operator Variability for Selected Methods for Measuring Bulk Specific Gravity of Hot-Mix Asphalt Concrete", *Transportation Research Record #1761*, Transportation Research Board, Washington, D.C., 2001.

11. Cooley, L.A. Jr., Prowell, B.D., Hainin, M.R., Buchanan, M.S., and Harrington, J., "Bulk Specific Gravity Round-Robin Using the CoreLok Vacuum Sealing Device", *NCAT Report No. 02-11*, FHWA-IF-02-044, National Center for Asphalt Technology, Auburn, Alabama, November 2002.

12. Spellerberg, P., and Savage, D., "An Investigation of the Cause of Variation in HMA Bulk Specific Gravity Test Results Using Non-absorptive Aggregates", Prepared for the National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C., July 2004.