Technical Note 0302

Assessment of Field Asphalt Density Gauge Data When Compared With Cores Processed According to AASHTO T-166

Prepared by:
TransTech Systems, Inc.
1594 State St.
Schenectady, NY 12304

March 24, 2004

Rev 03

INTRODUCTION

Gauge Capability Studies are generally conducted using guidelines such as described in ASTM E691 (Reference 1). In such studies sample data are specified to cover the complete range of the process being monitored. In a properly conducted Gauge Capability Study, factors that can affect bias, repeatability, reproducibility, and stability are carefully controlled to determine the influence of each on the overall measurement process. Each of these parameters are evaluated in a deterministic sequence that is designed to isolate the effects of each. Defined processing of the parameters determines the effect of each on the overall measurement. The processing determines whether any of the factors prevent the gauge from meeting overall accuracy requirements when measuring a specified process, such as asphalt paving,

For the case of paving QC/QA, in-place asphalt density is the acknowledged control parameter. The established direct measurement standard is the in-place core taken after completion of finish rolling in accordance with AASHTO T-166 (Reference 2). This process establishes density directly by determination of the mass in the volume occupied by the core by water displacement. The precision of T-166 is specified as +/-0.02 g/cm^3 = +/-1.25 pcf (Reference 2). System level influences such as operators, the coring process, and the specific process in T-166 add additional and unspecified reproducibility errors. Several gauges are available that measure asphalt density indirectly. Nuclear gauges measure scattering cross section and impedance gauges measure dielectric permittivity. Each technology can produce measurements that are proportional to the density of asphalt pavement. For best results on a given mat, the gauges are offset to match cores extracted from the same mat.

The total range of density possible in the paving process ranges from ~80% of MTD behind the paver to 92-96% of MTD after finish rolling. This represents a density range of 128-153 pcf if MTD is considered to be 160 pcf. QA/QC specifications typically call for determination of density to +/- 1.5pcf. The typical range of density values observed at different locations on the mat after finish rolling is only 4-8 pcf.

Determination of ground truth accuracy of pavement density is generally not possible as there is no absolute standard. AASHTO T-166 is the accepted standard by which all other gauges are judged. As absolute accuracy cannot be determined, many have attempted to compare gauges to AASHTO T-166 results using simple correlation. Current research (NCHRP Web Document 54 (Project D9-26): Contractor’s Interim Report) has shown that the 95% confidence values for the T-166 process range from 1.4-2.6 pcf. Many previous studies have assumed that the T-166 process is perfect, even though currently T-166 requires rejection of companion values that differ by more than 1.25 pcf. In the procedure proposed herein, the precision of the core process will be included in the analysis of gauge data to enable better statistical conclusions to be drawn, and to provide a recommendation for the number of cores required.

Further guidelines and suggestions for the design and data analysis appropriate for gauge comparison studies is provided in Appendix B.

Data Acquisition Procedure

There are two primary activities comprising the data acquisition activity. The first activity calibrates the instruments to the mat/mix by adjusting the offsets of the gauges so that the mean of a set of reference gauge and core readings are the same. The second activity involves taking the actual core and gauge data.

Reference Data Collection

1. Mark locations for a reference core and a minimum of two companions in a location that will be representative of the overall job. Insure the mat surface where measurements will be made is flat and free of surface indentations, roller marks, loose stone, etc. If the instrument rocks on the mat the surface is not flat enough.
2. Take 5 readings with each gauge at the location of the marked core circles. The PQI readings should be taken at the center and at equidistant locations on the circumference of a 3 in. circle centered on the core location. The PQI should be operated in continuous mode for this critical step so that a single bad reading does not compromise the offset operation to be conducted in Step 5. The nuclear gauge readings could be taken by rotation of the gauge in 5 equal angles with the center of the measuring volume centered on the core location. The important consideration is that the sensitive measuring volume of the instrument must coincide with the core location to minimize the effects of any segregation or other non-uniformities in the mat.
3. If the reference core results will be available prior to the actual paving, take the mean and standard deviation of the cores. If any of the cores is more than 2 standard deviations from the mean, do not include that core in the mean. *Note if the core data will not be available prior to the paving, then the data will be post corrected to the core data.*
4. Take the average of the PQI and nuclear gauge readings from all cores that were acceptable in Step 3. If any individual readings are greater than 2 standard deviations from the mean, reject those readings and redo the statistics.
5. Adjust the PQI and nuclear gauge offsets by the difference between the averages obtained in Steps 4 and 3. Use these settings for all data to be taken during the test.

Test Data Collection

1. During the test, it is beneficial if possible to take at least one companion for each test core. If any process variables change during the test, such as mix, mat thickness, base or top mat, identify each group of data accordingly. This will help isolate any interaction effects in the data analysis. It will also be helpful in the analysis if a set of gauge readings be taken behind the paver prior to rolling.
2. Take 5 PQI and nuclear gauge readings as before prior to removing each core. For each PQI reading record density, H_2O, millivolts, phase, and temperature. If any readings are to be taken in the rolling pattern, allow sufficient time for any surface water to evaporate (H_2O reading less than 10) prior to taking readings.

Data Analysis Procedure

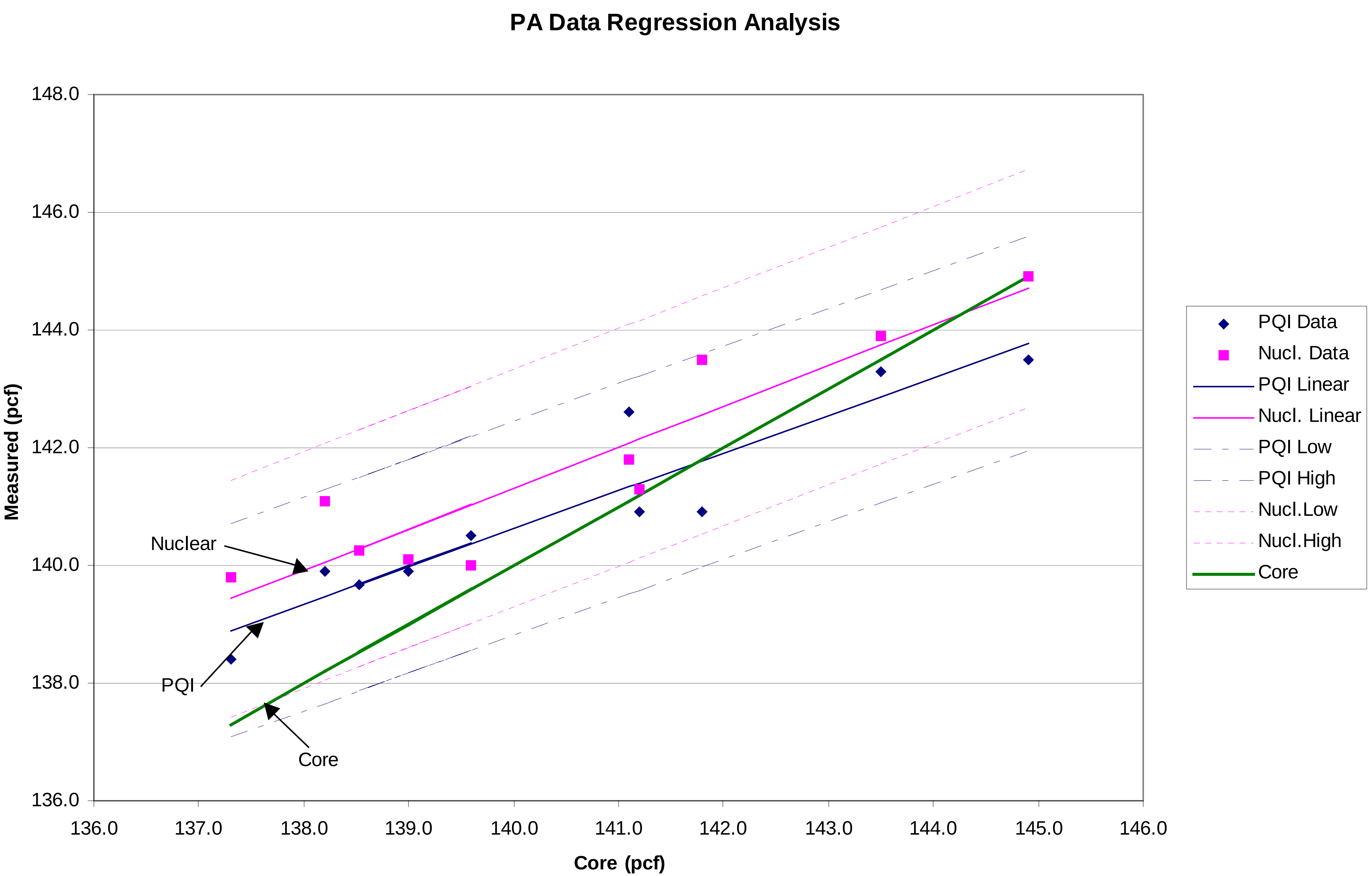
This analysis assumes that all data points, including the reference points taken in Steps 1&2 above form a uniform population. If any parameters, such as mix design, aggregate source, top or base, or mat thickness, changed during the test, each homogeneous group should be treated as a separate population for analysis purposes. Within each population, at least 10 sets of core/gauge readings should be taken.

In this analysis, the core readings will be used as the reference values. However adjustments will be made to the analysis to account for the fact that core process as defined by AASHTO T-166 is believed to have a 95% confidence reproducibility of 1.4-2.3 pcf depending on mix aggregate size. Bias of the core process is unknown as no absolute standard exists. For the purposes of this analysis the T166 bias will be assumed to be 0.

1. **Offset adjustment.** Insure that all gauge readings have been offset so that the reference means match the mean of the reference cores. (Step 5 above).
2. **Sort Datasets.** Sort the datasets so that the core readings are in ascending order.
3. **Correlation Analysis.** A linear correlation analysis will be performed on each population to determine if a *statistically significant linear relationship* exists between the gauge data and the core data. Using the CORREL operator in EXCEL determine the correlation of each gauge reading set with the corresponding core readings. For each population use the formula in Appendix A to determine with 95% confidence that a relationship exists for the actual number of data points in the population. Any data set not meeting the 95% test will not be analyzed further as no strong statistically significant correlation exists for that data set. *(Note that outliers can significantly affect the results of the correlation operation. If a dataset is rejected by this test continue with Step 3. to determine if the rejection was due to an outlier(s).) Also note that relative correlation values cannot be used to make a determination that one instrument is better than another.)*
4. **Outlier Identification.** For datasets with relatively small populations, results are strongly affected by outliers. This procedure will identify points that are outliers to 95% confidence. Using the EXCEL LINEST function perform a linear regression on each gauge dataset using the core data as the independent variable. LINEST determines the linear fit parameters, the error estimates for each parameter, the standard error of the fit (SEy), the R^2 value, the F-test value, and the degrees of freedom (df). As LINEST assumes that there is no error in the independent variable an adjustment to SEy will be made to account for the known variability of the core process. The effective SEy (at 95% confidence) will be given by the RSS of the 2 sigma values of the regression SEy and the coring process. For the core process, the 2 sigma values will be 1.4 pcf for a 12.5mm mix and 2.3 pcf for a 19mm mix. Now each datapoint will be compared with the effective SEy at 95% confidence. Any datapoint for which the magnitude of $(x-x_{fit})>SEy_{eff}$ will be rejected. If any points are rejected, the analysis for that dataset will be repeated from Step 3 without the outliers.
5. **Regression Significance Test**. The F-test will be performed to determine whether the regression fit is statistically significant in explaining the data. The value of F-critical will be looked up in a standard table for 95% confidence and the number of datapoints in each population. The F values for each regression calculated by LINEST will be compared with the F-critical value. F values larger than F-critical indicates that the regression is statistically significant and there is a 95% confidence that chance did not produce the apparent relationship. For any dataset for which the value of F is less than F-critical no

conclusions can be drawn. (*Note that relative F-test values cannot be used to make a determination that one instrument is better than another.)*

6. **Graphical Analysis**. Generate an XY scatter plot with the Core data as the independent variable. Plot the following dependent variables. Core (as line), and for each gauge, the raw data points, regression line, and the upper and lower 95% confidence lines (from the 95% confidence SEy_{eff} calculated in Step 4). Figure 1 shows a typical plot for a test conducted on finished asphalt. Consider the highest upper 95% and the lower 95% curves as the statistical boundaries for each instrument for this test. If the core line curve is outside the instrument bounds curve for either instrument, a statistically significant difference exists between the cores and the corresponding instrument. This difference must result from bias or linearity difference between the cores and the instrument as repeatability has been incorporated into the confidence limits. If the upper and lower confidence limits from one instrument are within the limits of the other instrument and the core line plot is within the instrument 95% limits then the only conclusion that may be drawn is that that both instruments are equally good in measurement of density for the test conditions.


PA Data Regression Analysis
148.0
146.0
144.0
142.0
140.0
138.0
136.0
Measured (pcf)
136.0
137.0
138.0
139.0
140.0
141.0
142.0
143.0
144.0
145.0
146.0
Core (pcf)
Nuclear
PQI
Core
PQI Data
Nucl. Data
PQI Linear
Nucl. Linear
PQI Low
PQI High
Nucl.Low
Nucl.High
Core

Appendix A – Significance Test for Linear Correlation

Without entering into details, we will test if r is significantly different than 0 by building an interval of confidence around r. If 0 is included in the interval, r will **not be considered as significantly different than 0** and if 0 is not included in the interval, r will **be considered as significantly different than 0**.

To build that interval, we will use a Fisher's Z-transformation (1.96 in formula means that we work at the 95% confidence level).

The formula is:

$$\frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \pm 1.96 * \sqrt{\frac{1}{n-3}} = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

with

n the sample size.

r the correlation coefficient.

ρ the high end and low end case of the interval.

If 0 is not between the 2 ρ, r is significantly different than 0

For example for a data set with n =10 and r = 0.7

C1 = 0.5*ln((1+0.7)/(1-0.7)+1.96*sqrt(1/(7) = 0.867+0.74 = 1.607

C2 = 0.5*ln((1+0.7)/(1-0.7)-1.96*sqrt(1/(7) = 0.867-0.74 = 0.127

For this set 0 is not in the interval between C1 and C2 so significance is established at the 95% level.

Appendix B -- General Suggestions on Gauge Study Design and Data Analysis

Many studies comparing methods for measurement of asphalt density or related quantities (such as % air voids) are conducted using measurements made on finished courses. Such data will exhibit a narrow span of density, generally 3-5 pcf. As the 95% confidence band of AASTO T-166 and many measurement devices is of similar order (1.5-2.5 pcf) it is very difficult to generate statistically significant conclusions with small sample sets. There are only two remedies to this difficulty; data sets with span much greater than the gauge noise band or a large enough sample set to reduce the effective gauge noise band well below the span of the data. For example for a 25mm course the published D2S for AASTO T-166 (from Reference 1) is 2.4 pcf. Lets say the gauge D2S is 2.5 pcf. The RSS of the two yields 3.5 pcf as the joint noise confidence band at 95% confidence. Let's further say that the span of the data is 5 pcf. In order for the gauge noise band not to affect the conclusions, the effective gauge noise band should be less than 10% of the span. If the gauge noise is normally distributed, then the effective noise band is reduced by the square root of the number of data points. Therefore 76 data samples would be required to reduce the measurement noise band below 10% of the span of the data.

Small sample populations (<10) are also very sensitive to outliers in the datasets. As a result, an analysis must be conducted on each dataset to identify and remove outliers prior to any further processing. This can be accomplished by performing a regression of each data set against the standard and rejecting any samples for which any of the elements are outside the joint 95% noise band. All elements must be rejected, not just the element outside the noise band as the bad data point could be core value and not one of the instrument readings. The outlier analysis serves to identify that members of a sample set are not consistent with each other; it does not identify which is bad in some absolute sense. Reference 5 provides a sample of the analysis procedure illustrating the points made above.

A final note is a caution on the use of correlation analysis alone to compare measurement data. The correlation coefficient (as produced e.g. by the EXCEL function CORREL) measures the strength of a linear relation between two variables, not the agreement between them. There is perfect agreement only if all the points lie on the line of equality, but there is perfect correlation if the points lie on any straight line. Thus offset and scale factor differences between two data sets do not affect the correlation but certainly affect the agreement. The correlation coefficient is influenced by the span of the data; the wider that span the higher the correlation coefficient. This artifact of the correlation operation has nothing to do with the agreement of the data. For small datasets, an apparently high correlation coefficient may be due to chance. Other statistical tests, such as the F-test, can be used to establish the statistical significance of a relationship given the desired confidence level, number of variables, and the sample size. The correlation coefficient itself, when properly used, must be tested against the c-critical value published in statistical tables that give the minimum value of the correlation coefficient that can be considered to indicate that a statistically significant linear relationship exists between two data sets given the desired confidence level and the number of data points.

Appendix C -- References

1. ASTM Standard E691-99, "Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method".
2. Spellerberg, P. et al, "Precision Estimates of Selected Volumetric Properties of HMA using Non-Absorptive Aggregate", NCHRP Web Document 54 (Project D9-26), dated February 2003.
3. ISO/IEC 17025:2000 STANDARD "General requirements for the competence of testing and calibration laboratories".
4. ANSI/NCSL Z540-2-1997 "American National Standard for Calibration – US Guide to the Expression of Uncertainty in Measurement".
5. Spreadsheet implementing the procedure in Reference 5 for a data set comparing two gauges against the core process on finish data.