Evaluation of Rutting Variations in FHWA Full-Scale Accelerated Pavement Testing

X. Qi
Turner-Fairbank Highway Research Center, 6300 Georgetown Pike, McLean, VA 22101 USA

T. Mitchell
FHWA, 6300 Georgetown Pike, McLean, VA 22101 USA

ABSTRACT: The Federal Highway Administration (FHWA) has been conducting full-scale pavement tests using two Accelerated Loading Facility (ALF) machines for over 18 years. Due to the expense of full-scale pavement testing, most pavements were tested without replication. Variation in performance in replicate tests becomes a major concern when aging takes place in the asphalt pavements. In order to increase our confidence in full-scale testing results, replicate tests were conducted during rutting experiments in a recently completed Superpave validation study (1993-2001). This paper summarizes the results of replicate tests, which were conducted on 12 pavement test sites. Statistical tests were first performed on the rutting results in the replicate pavements. Asphalt aging effect on the rutting was discussed. The variation components were identified and accuracy curves were developed for all test pavements. Finally, the rutting variations among the different pavement layers were also evaluated.

KEY WORDS: APT, asphalt pavement, rutting, variation.

1 INTRODUCTION

In 1986, the Federal Highway Administration (FHWA) established its Pavement Testing Facility (PTF) at the Turner-Fairbank Highway Research Center (TFHRC) in Virginia. Since then, several research studies have been conducted at the FHWA's PTF. The first research study was directed at establishing equivalency factors, to relate the results of Accelerated Loading Facility (ALF) machine tests to rutting under actual truck traffic. The effects of increased tire pressure on flexible pavement response were also evaluated during this initial study (Bonaquist et al. 1989). The next PTF study involved assessing the impact of wide-based single tires on flexible pavement rutting and cracking performance (Bonaquist 1992). Next an ALF was transported to several western states, for field tests on in-service highways. From 1993 to 2001, the FHWA's ALF work focused on the validation of the Superpave specification for asphalt binders (Sherwood et al. 1998 and 1999). From 1998 to 2000, the ALF machines were also used for full-scale accelerated testing of Ultra-Thin White (UTW) Topping (64 and 89 mm thicknesses of Portland cement concrete over hot-mix asphalt pavement) study (Qi et al. 2004). The current ALF project is aimed at the development of an improved Superpave binder specification (Mitchell et al. 2004).

Due to the expense of full-scale pavement testing, most pavements in the past were tested without replication. Variation in performance in replicate tests becomes a major concern when aging takes place in the asphalt pavements. In order to increase our confidence in full-scale testing results, replicate tests were conducted during rutting experiments in the Superpave validation study from 1993 to 2001. This paper summarizes the results obtained of replicate tests on 12 pavement test sites. Statistical tests were first performed on the rutting results in the replicate pavements. Asphalt aging effect on the rutting was discussed The variation components were identified and accuracy curves were developed for all test pavements. Finally, the rutting variations among the different pavement layers were also evaluated.

2 PAVEMENT TESTING FACILITY

The Pavement Testing Facility is a permanent, outdoor laboratory located on the grounds of the TFHRC. The facility currently includes two ALF machines and 12 full-scale pavement lanes, with 48 test sites available. Each ALF is a 29-m long frame with rails to direct rolling wheel loads. The wheel loads can range from 44 to 100 kN. Loads roll at 18.5 km/h (5.1 m/s) over a 14-m long by 1-m wide section of pavement. The center 10 m are used for data collection. An ALF loads the pavement in one direction, and the loads can be laterally distributed. In both the 1993-2001 study and the current program a super-single tire is being used. Prior ALF tests showed "super-single" tires cause about twice as much fatigue and rutting damage as do traditional dual tires. The tests are stopped at least once every three days, or 25,000 load repetitions, for routine maintenance and data collection. A radiant heat system was used to maintain a relatively constant pavement temperature, at or above ambient, during a single test; this temperature can be held as high as 76 °C in the summer and as low as 10 °C in the winter.

Conventional flexible pavement test sections were constructed using normal highway materials, equipment and procedures. The surface layers in the 1993-2001 study were constructed with five different asphalt binders and two different aggregate gradations . The base course for all of the test sections consisted of a dense-graded, crushed diabase meeting Virginia Department of Transportation (VDOT) gradation 21-A. The subgrade classifies as an A-4 silty clay. During construction, the upper 610 mm of the subgrade was removed and recompacted. A woven geotextile was placed over the completed subgrade prior to construction of the base course. Although the 12 test lanes in the 1993-2001 study included two different pavement cross sections, Lanes 9-12, the lanes of interest here, all have 200-mm hot-mix asphalt surface courses over a 460-mm base course.

Table 1 summarizes the lanes and sites used in the replicate testing. The two binders used were unmodified asphalts from Venezuela's Lagoven base stock. Two different dense-graded asphalt concrete mixtures were used in these lanes: a 19.0-mm nominal maximum size mixture meeting VDOT designation SM-3B was used in Lanes 9 and 10, and a 37.5-mm nominal maximum size mixture meeting a modification of VDOT designation BM-3 was used in Lanes 11 and 12. The 19.0-mm mixture is essentially a surface course mixture and was placed in 50-mm, compacted lifts, while the 37.5-mm mixture is typically a base course mixture and was placed in 100-mm, compacted lifts.

Each pavement lane was divided into four testing sites, which thus have the same materials and pavement structure. Table 1 also presents the assigned replicate testing sites and the testing date for each. Note that the number of replicate testing sites in each lane varied from 2 to 4, depending on the availability for the rutting experiment.

Table 1 Summary of Lanes and Sites Used in Replicate Testing

Lane Number	Binder Type	Performance Grade	Mix Agg. Gradation	Top Size Agg. (mm)	Testing Site	Testing Time
9	AC-5	PG 58-34	Surface	19	2	Jul-94
					1	Jul-95
					4	Jul-98
10	AC-20	PG 64-22	Surface	19	2	Jul-94
					1	Jul-95
					3	Jun-98
					4	Jul-98
11	AC-5	PG 58-34	Base	38	2	Jul-94
					1	Aug-95
					3	May-98
12	AC-20	PG 64-22	Base	38	1	Jul-94
					3	Sep-01

3 INSTRUMENTATION

Instrumentation to collect pavement temperature and performance data was installed in each site prior to testing. Figure 1 shows the layout of this instrumentation. Two sets of thermocouples were used to record the pavement temperature every 30 minutes at depths of 0, 20, 102 and 197 mm. Transverse and longitudinal profiles were obtained, after selected numbers of load repetitions, with a semiautomatic device. Eight transverse profiles were collected, at 1.22-m intervals. Longitudinal profiles were collected at the centerline of the wheel path. Asphalt layer rutting data were collected by performing differential surveys on a series of reference plates installed at the time of construction and also shown in Figure 1.

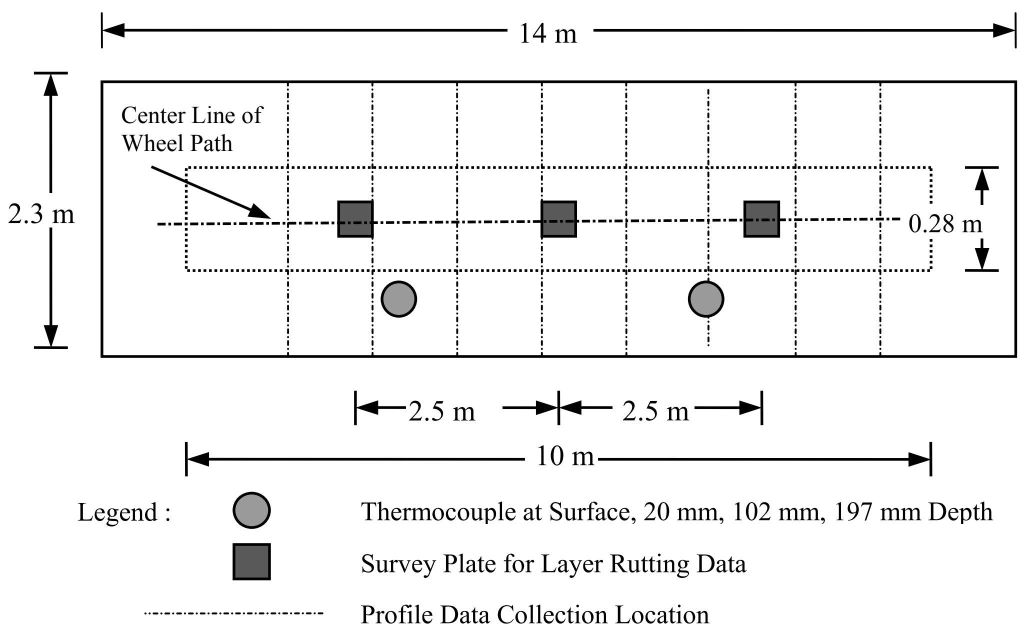


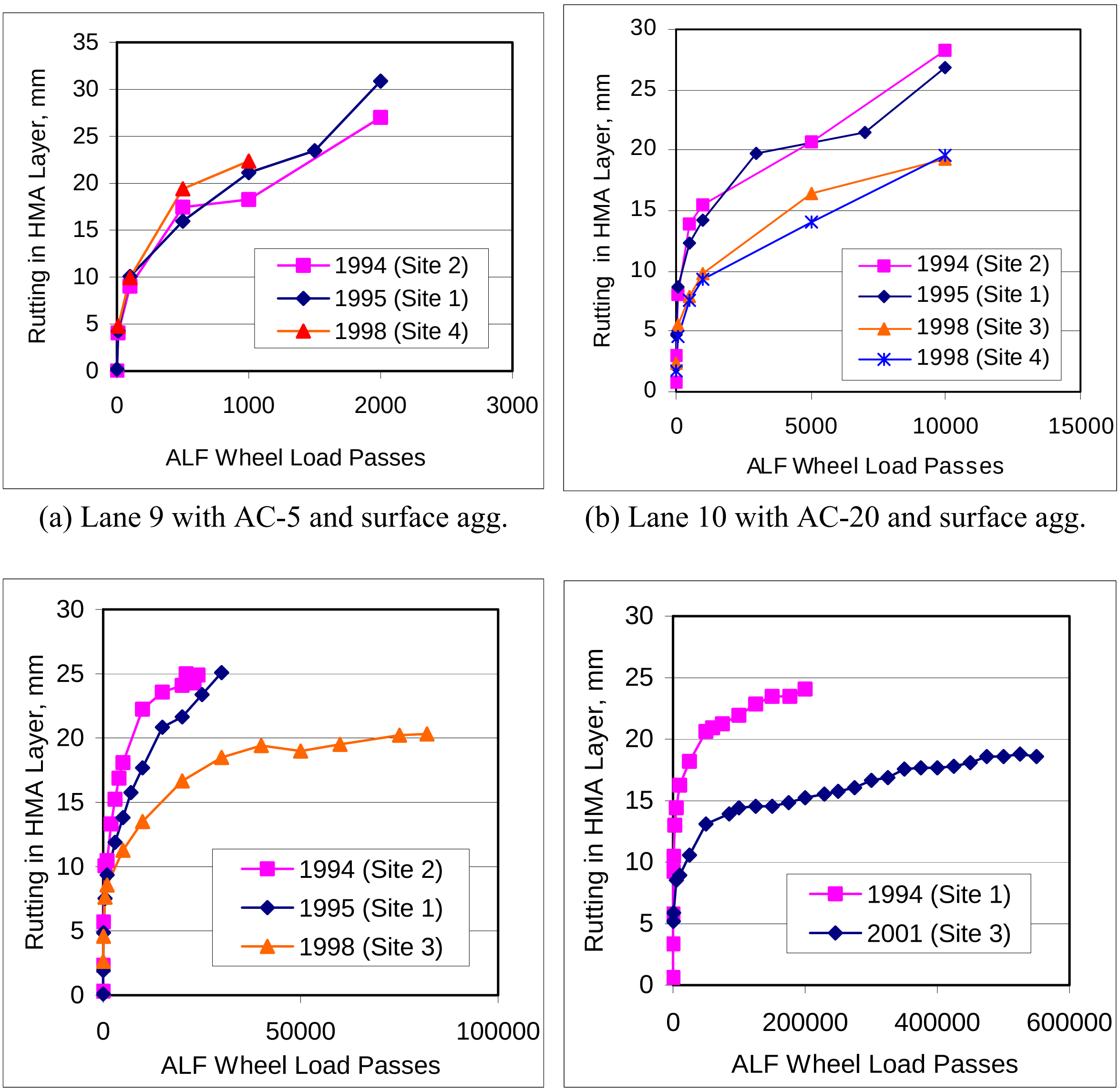
Figure 2 Typical instrumentation layout for a test site (not to scale)

The reference plates were located at the surface of the pavement and on top of the aggregate base. Three sets of reference plates were spaced equally along the centerline of the test section. Differential surveys on these plates measure the permanent displacements at the surface and at the top of the base. The difference between these two measurements yields the amount of rutting in the HMA layer alone.

4 TEST RESULTS AND EVALUATION

4.1 Replicate Testing Results

All the replicate tests in the rutting experiment were conducted at one temperature, 58 °C, and at a 44-kN wheel load without transverse wheel wander. Graphs of the rutting data for the asphalt layers are presented in Figure 2. Each data point in the figure is the mean of three measurements from the survey plates. Generally, the asphalt layer



(a) Lane 9 with AC-5 and surface agg.

(b) Lane 10 with AC-20 and surface agg.

(c) Lane 11 with AC-5 and base agg.

(b) Lane 12 with AC-20 and base agg.

Figure2: Comparisons of rutting in HMA layer at different pavement ages.

rutting among the replicate tests varied significantly, depending on the combinations of the binder types and aggregate sizes. Lane 9, with AC-5 binder and surface aggregate, showed the least variation in layer rutting among the replicate tests. Lane 10, with AC-20 binder and surface aggregate, showed a relatively small variations in two tests conducted in 1994 and 1995 and in two tests conducted in 1998, but also showed a significant difference between the two pairs. Lane 11, with AC-5 binder and base aggregate, showed progressively lower rutting rates as the pavement aged. Lane 12, with AC-20 binder and base aggregate, showed the highest variation between two rutting tests, but also had the longest time interval, 7 years, between two tests. Statistical tests on the mean values of asphalt layer rutting are reported in the next subsection.

4.2 Statistical tests on the rutting results

For each lane, the mean values of rutting in asphalt layer from each site at different common ALF loading passes, were tested by ANOVA to see if they are significantly different among the different sites. Comparing rutting on different lanes after an identical number of passes is a typical way of reporting accelerated pavement testing data. A 95 percent confidence interval was chosen for the analysis. The test results are showed in Table 2.

Table 2 Results of ANOVA Test for Asphalt Layer Rutting

Statistical analysis is performed at $\alpha = 0.05$,
i.e., at a 95 percent confidence level.
E = Statistically equal / not significantly different at $\alpha = 0.05$
N = Statistically not equal / significantly different at $\alpha = 0.05$

Lane 9		Lane 10		Lane 11		Lane 12	
ALF Pass	Result	ALF Pass	Result	ALF Pass	Result	ALF Pass	Result
10	E	10	N	10	E	500	N
100	E	100	N	100	N	1,000	N
500	E	500	N	500	E	5,000	N
1,000	E	1,000	N	1,000	E	10,000	N
2,000	E	5,000	N	5,000	N	50,000	N
		10,000	N	10,000	E	100,000	N
				20,000	E	200,000	N

As seen from the table, for Lane 9 the ANOVA test results showed all E's at the listed ALF load passes, which indicates that the mean values of rutting from each site are statistically identical, i.e., not significantly different at the $\alpha = 0.05$ level, although these sites were tested at different times. On the other hand, for Lanes 10 and 12, the ANOVA test results showed all N's at the listed passes, which indicates that at least one pair of test sites have yielded significantly different rutting histories in the asphalt layer. Further t-tests were performed on Lane 10 rutting data at the listed loading passes to see which pair of test sites are significantly different. Again, a 95 percent confidence interval was chosen for the analysis. No significant difference was found between results of the two tests in 1994 and 1995, or between the two tests in 1998. However, the rutting results obtained in 1998 are significantly different from those obtained in 1994 and 1995. This may relate to asphalt binder aging and will be discussed further in the next subsection. Since Lane 12 only had two test sites, no further t-test is needed; the ANOVA test shows the mean values of rutting in the 2001 test on this lane are significantly different from those in the 1994 test. For Lane 11, ANOVA test results showed five E's and two N's at different ALF load passes. A close examination of the ANOVA results reveals that the calculated F values are very close to the critical F values for the two N cases, which indicates the differences for these two cases are marginal.

4.3 Asphalt Aging Effect

In order to assess the binder aging effect, pavement cores were taken at the end of the rutting tests on Lanes 9, 10, and 11. Dynamic shear rheometer (DSR) tests were performed on the binders recovered from the pavement cores. The $G^*/\sin\delta$'s at 58 ^oC and DSR frequency 2.25 rad/s are summarized in Table 3. Comparisons of the 1994 and 1998 binder data indicate that significant binder aging or hardening has taken place in the Lanes 10 and 11 but not in Lane 9. This may explain why three replicate rutting tests in Lane 9 are statistically equal, while the 1998 tests in Lane 10 are statistically different from the 1994 tests. However, in Lane 11, the binder has hardened (according to the DSR test) but this is not reflected in the rutting results; this may be attributable to the larger size of aggregate used in Lane 11, i.e., the influence of aggregate on the rutting was much more significant than that of AC-5 binder.

Table 3 $G^*/\sin\delta$'s at different ages

Lane	Mixture	$G^*/\sin\delta$ of the neat binders after RTFO[+], kPa	$G^*/\sin\delta$ of the recovered binders, kPa	
			1994	1998
9	AC-5 Surface	0.66	1.3	1.4
10	AC-20 Surface	2.7	4.3	10.4
11	AC-5 Base	0.66	1.7	6.4

[+]Rolling Thin Film Oven

Due to the asphalt aging effect, the rutting on Lane 10 tested in 1998 was significantly reduced in comparison to the replicate test in 1994. It is necessary to consider such aging related reduction in comparison of rutting results tested at different ages. One way to achieve this is to predict pavement rutting with the binder parameter, $G^*/\sin\delta$ incorporated into the prediction model. As long as how much change in the $G^*/\sin\delta$ caused by the asphalt aging is known, the rutting reduction then can be predicted. Sherwood and Qi have related permanent deformation model parameters, μ and α to asphalt binder parameter $G^*/\sin\delta$ for both modified and unmodified binders based upon the ALF pavement calibration in this Superpave Validation study (Sherwood et al, 1998). These relationships and the rutting prediction model are presented as follows:

$$\alpha_{unmodified} = 0.6483 + 0.0929 \ln (G^*/\sin\delta) \quad (1.a)$$

$$\alpha_{modified} = 0.7729 + 0.0567 \ln (G^*/\sin\delta), \quad (1.b)$$

$$\mu_{all} = 0.5083 + 0.2894 \ln (G^*/\sin\delta), \quad (1.c)$$

$$RD = H\,\varepsilon_p = H\,\varepsilon_r \frac{\mu}{1-\alpha} N^{1-\alpha} = D \frac{\mu}{1-\alpha} N^{1-\alpha} \qquad (1d)$$

Where

RD = rut depth within the HMA layer
H = Layer thickness
ε_p = permanent strain
ε_r = dynamic strain
N = number of load repetitions
μ, α = model parameters
D = layer deflection

Note that the above relationships between model parameter and binder parameter were developed only for the mixtures with the surface course aggregate in this project. By using these relations and the rutting prediction model, the predicted rut depths in HMA layer of Lane 10 are compared with the measured values (Figure 3). Although the model over predicted slightly, it provided a useful tool for estimating the effect of aging on rutting.

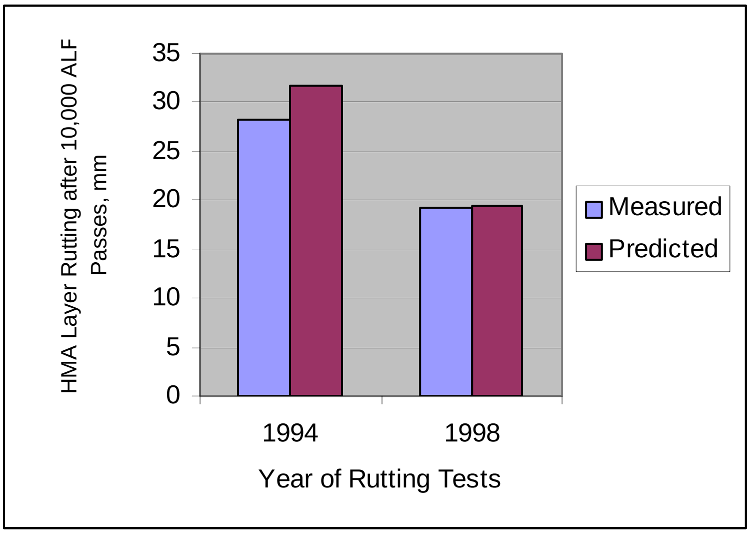


Figure 3. Comparison of predicted and measured rutting for Lane 10

4.4 Variance Components

Since Lane 10 had replicate tests at different ages, it was used for further analysis of variance components. The total variance can be separated into different components and be mathematically expressed as follows:

$$\sigma_T^2 = \sigma_{within\ site}^2 + \sigma_{between\ site}^2 + \sigma_{aging}^2 \qquad (2)$$

where σ_T^2 = total variance in asphalt layer rutting from four sites at a given number of ALF passes.
$\sigma_{within\ site}^2$ = variance within a site
$\sigma_{between\ site}^2$ = variance between sites
σ_{aging}^2 = variance associated with the binder aging

With the analysis of variance table, the variance components can be calculated (Box et al., 1978), and they are summarized in Table 4. Columns 2 to 4 show the variance components, while columns 5 to 7 show the ratio of component variance to total variance in percent. At all ALF load passes, the aging components are dominant, followed by variance within site. The variance between sites is the smallest among the three components. The last row of the table shows the average values of variance components over the different passes. Therefore, in average, 80 percent of the variation is associated with aging, 14 percent is within site, and 6 percent is between sites.

Table 4 Variance components of asphalt layer rutting for Lane 10 four sites

ALF	Variance Components, mm sq.			Component to Total Variance (%)		
Loads	Aging	Between Sites	Within Site	Aging	Between Sites	Within Site
10	0.047549	0.025197	0.034138	44	24	32
100	0.222098	0.001930	0.029870	87	1	12
500	0.547014	0.016053	0.094488	83	2	14
1000	0.725424	0.016866	0.067666	90	2	8
5000	0.848563	0.083210	0.131978	80	8	12
10000	1.281227	0.006083	0.048578	96	0	4
Mean	0.611979	0.024890	0.067786	80	6	14

4.5 Rutting Measurement Accuracy Curves

In order to establish the accuracy curves (confidence intervals) of the rutting measurements from three survey reference plates, coefficients of variation (CV) were calculated for each set of rutting measurements using all available data in this Superpave validation project. Figure 4 is a plot of CV values versus mean asphalt layer rutting. The CV values are relatively constant across the mean values of asphalt layer rutting. Similar trends were observed for surface (total) rutting and for top of base rutting although they are not shown in this paper. Therefore the grand mean CV values of 18, 23, and 38 from surface, asphalt layer, and top of base were used to develop the rutting measurement accuracy curves. Figure 5 shows these curves in terms of relative error versus sample size (the number of plates). The figure clearly indicates that three measurements from plates are in the sensitive zone of the curves, i.e., an increase in sample size could significantly reduce the error. Therefore, in the PTF project initiated in 2002, the reference survey plates were increased from three to eight therein each test site. According to the curves, the relative error in asphalt layer rutting measurement will be reduced from 27 percent to 17 percent in this new project.

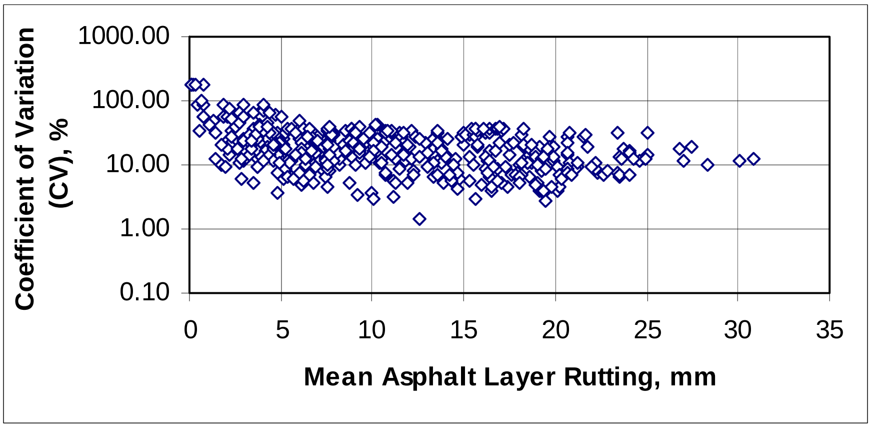


Figure 4 Coefficient of variation vs. mean asphalt layer rutting

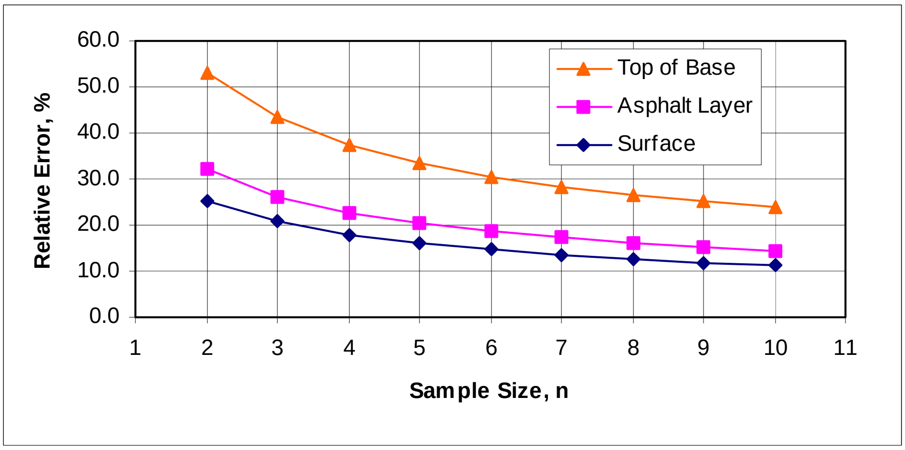


Figure 5 Rutting measurement accuracy curves

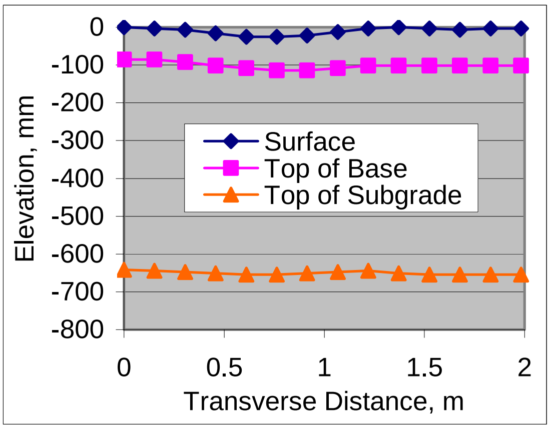
4.6 Layer Rutting Evaluation

At the end of ALF testing, test sites were trenched to inspect any rutting occurred in the subgrade soil. Four test sites were selected for trench study: Sites 3 and 4 in Lane1 and Sites 3 and 4 in Lane 5. Lane 1 was constructed with 100 mm of 19.0-mm HMA over 560 mm of base course, and Lane 5 was constructed with 200 mm of 19.0-mm HMA over 460 mm of base course. Figure 6 shows photos and measured profiles from the trenches for Lane 1 Site 4 and Lane 5 Site 4. The Lane 1 site was actually used for a fatigue test while the Lane 5 site was used for a rutting test. Inspection of the trenches and examination of the measured profiles from all four test sites indicated that little rutting occurred in the subgrade soil. Therefore, the rutting measured from the plates at the top of the aggregate base represented the aggregate base layer rutting and the rutting measured at the surface represented the total rutting of both asphalt and aggregate base layers.

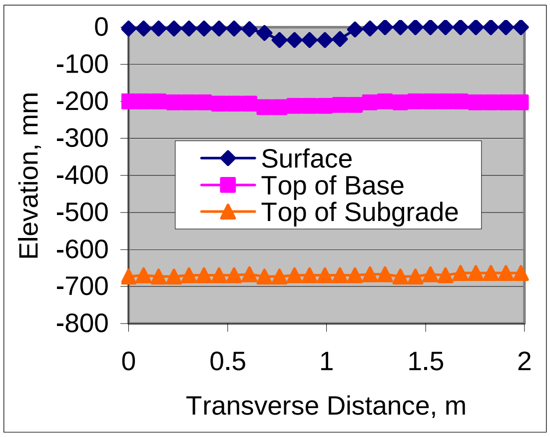
With the rutting measurements available for both layers, the rutting proportion in these two layers can be evaluated. The ratio of asphalt layer rutting to total rutting was used to evaluate the rutting proportion. A high value of this ratio (approaching 1.0) indicates a high portion of the rutting in the asphalt layer and a lower portion in the aggregate base layer. Figure 7 shows the histogram of the ratio for both rutting tests and fatigue tests, conducted in this Superpave

(a) Cross section of Lane 1 Site 4

(b) Cross section of Lane 5 Site 4



(C) Measured Profiles of Lane 1 Site 4



(d) Measured profiles of Lane 5 Site 4

Figure 6 Photos and profiles from trenches

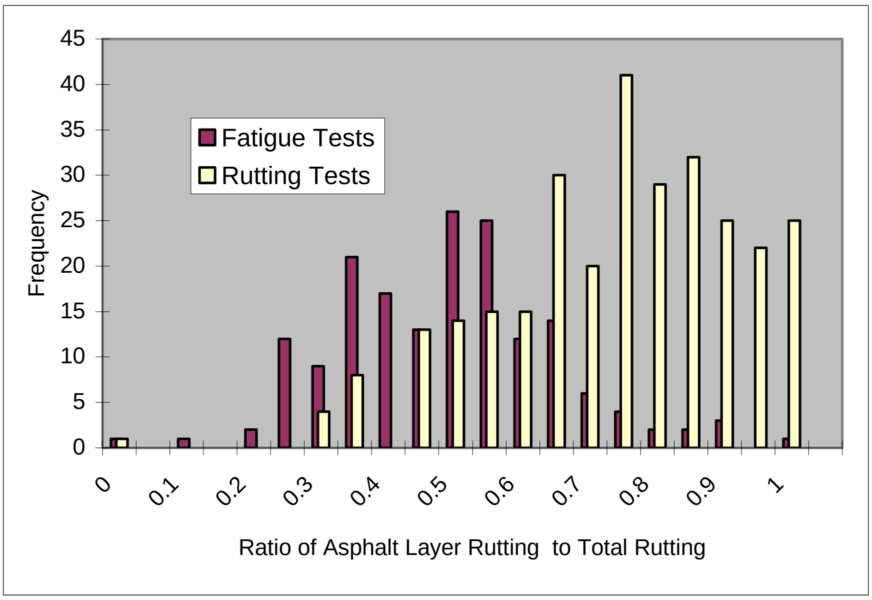


Figure 7 Histogram of rutting ratio

validation project. The histogram summarizes 295 rutting measurements, with a mean value of 0.72 and a standard deviation of 0.18, and 171 fatigue measurements, with a mean value of 0.46 and a standard deviation of 0.17. Clearly a greater proportion of the rutting occurred in the asphalt layer in rutting tests than in fatigue tests. Note that the rutting tests were conducted at high temperatures (52 to 76 ^{o}C), while the fatigue tests were conducted at intermediate temperatures (10 to 28 ^{o}C). On average, some 72 percent of rutting occurred in the asphalt layer during the rutting tests, while about 46% of rutting occurred in the asphalt layer during the fatigue tests.

5 SUMMARY AND CONCLUSIONS

- A total of twelve test sites in four pavement lanes, with two different binders and two different aggregate maximum sizes, were included in the rutting replicate tests at different times after construction. The rate of development of asphalt layer rutting among replicate tests at different ages varied significantly, depending on the combinations of the binder types and aggregate sizes.
- For Lane 9, with AC-5 binder and surface course aggregate, the mean values of asphalt layer rutting from three replicate tests in 1994, 1995, and 1998, respectively, were not significantly different statistically at all recorded ALF loading passes.
- For Lane 10, with AC-20 binder (much stiffer than AC-5) and surface course aggregate, the asphalt layer rutting from two replicate tests in 1998 were significantly different from those from replicate tests during 1994 and 1995. However, no significant difference was found within each pair, i.e., between the 1994 and 1995 tests, and between the two tests in 1998.
- For Lane 11, with AC-5 binder and base course aggregate (much larger aggregate size than surface aggregate), the mean values of asphalt layer rutting from three replicate tests, were statistically equal at most ALF loading passes, and marginally different at some loading passes.
- For Lane 12, with AC-20 binder and base course aggregate, the mean values of asphalt layer rutting in the 2001 test were significantly different from those in 1994 tests at all ALF loading passes.
- The binder rutting parameter $G^*/\sin\delta$, obtained from DSR tests on binders recovered from pavement cores and cut after the rutting tests in 1994 and 1998, indicates that significant aging or hardening of the binders had occurred in Lanes 10 and 11, but not in Lane 9. This may explain why the replicate results from Lane 9 rutting tests were statistically equal, but cannot explain the Lane 11 results where large aggregate may play an important role.
- Analysis of variance components on Lane 10 indicates that most of the asphalt layer rutting variation in this lane is associated with aging, followed by the variations within site and between sites.
- Rutting measurement accuracy curves were established based on all available rutting measurements. The use of three reference survey plates for rut determinations in a single site were found inadequate, and the use of eight plates was recommended for the rutting measurements in each test site in the 2002 new project.
- A trench study conducted at the end of ALF testing showed little rutting in the subgrade soil. An evaluation of the ratio of asphalt layer rutting to total rutting indicates that about 72 percent of total rutting occurred in the asphalt layer during the rutting tests (at a higher range of temperatures) while about 46 percent of rutting occurred in the asphalt layer during fatigue tests at an intermediate range of temperatures.

6 RECOMMENDATION

- Based on the evaluation of the rutting replicate results, no significant difference exists in the HMA layer rutting between any pair of replicate tests conducted in 1994 and 1995 (about a one year period). However, asphalt binder aging did cause a significant difference in replicate test results in a period of 3 to 4 years for a typical surface course mixture used in this project. Therefore, it is highly recommended that comparative APT rutting tests be completed within one-year. If the APT rutting testing has to last several years, aging effects should be considered through appropriate pavement modeling.

REFERENCES

Bonaquist, R., Surdahl, R., and Mogawer, W., 1989. *Pavement Testing Facility – Effect of Tire Pressure on Flexible Pavement Response and Performance*. FHWA-RD-89-123, Federal Highway Administration, Department of Transportation, Washington, D.C..

Bonaquist, R., 1992. *An Assessment of the Increased Damage Potential of Wide-Based Single Tires*, Proceedings of 7th International Conference on Asphalt Pavements, Nottingham, England.

Box, G., Hunter, W., and Hunter, J., 1978. *Statistics for Experimenters, An Introduction to Design, Data analysis, and Model Building*, John Wiley & Sons, New York.

Mitchell, T., Stuart, K., Qi, X., Al-Khateeb, G., Youtcheff, J., and Harman, T., 2004. *ALF Testing for Development of Improved Superpave Binder Specification,* Proceedings of 2nd International Conference on Accelerated Pavement Testing, Minneapolis, Minnesota.

Qi, X., Mitchell, T., and Sherwood J.A., 2004 *Evaluation of UTW Fatigue Cracking Using FHWA's Accelerated Loading Facility*, Proceedings of the Fifth RILEM International Conference on Cracking in Pavements, Limoges, France.

Sherwood, J. A., Thomas, N. L., and Qi, X.,1998. *Correlation of Superpave* $G^*/\sin\delta$ *with ALF Rutting Test Results*, Transportation Research Record 1630, Asphalt Mixtures Stiffness Characterization, Variables, and Performance, Transportation Research Board, Washington, D.C..

Sherwood, J. A., Qi, X., Romero, P., Stuart, K.D., and Naga, S., 1999. *Full-Scale Pavement Fatigue Testing from FHWA Superpave Validation Study*, Proceedings CD-ROM of 1999 International Conference on Accelerated Pavement Testing, Reno, Nevada.