The reliability of asset management regime of the SROH using air void content of asphalt mixtures

Corresponding Author: Dr. Monower Sadique**,** Lecturer, Department of Civil Engineering,
University of Bolton, Bolton, BL3 5AB, UK.
Email: sadiquerhd@hotmail.com ; m.sadique@bolton.ac.uk

Hassan Al-Nageim, Professor of Structural Engineering, School of Built Environment,
Liverpool John Moores University, Email: H.K.Alnageim@ljmu.ac.uk

Karl Stopps MInstRE, Head of Street Works, Carillion Telent, UK.
Email: karl.stopps@carillionplc.com

Abstract

The comparison pairs of cores (each 100mm apart) from 68 reinstatements sites from various parts of the UK revealed that, the compounding consequences of generic non-homogeneous characteristics of hand laid recipe mixed materials (specified in SROH) and high likeliness of being biased during AV testing makes the coring method extremely unreliable with very low repeatability and reproducibility. The wide-ranging maximum density reported in every instance in the comparison pair coring experiments meaningfully rationalizes the distorted homogeneity of materials. Although not only maximum density but also bulk density of adjacent cores located only 100 mm apart were found to be varied in the case of every pair in this study. Furthermore, the in-situ performance shown by from 50 reinstatements after experiencing 1.5 years to 10 years real life aging from various parts of the UK predictably indicates that either the linkage between the reinstatement with non-compliant AV and its impact on footways durability is non-proven or the reported AV content is extremely over estimated. At 95% level of significance, there exists enough evidence to conclude that, due to high uncertainty, very low repeatability and reproducibility and poor reliability with high chances of bias, the assessment of hand laid reinstatement work by air void (AV) testing will expose both the contractor and the client to unacceptable risk.

Key words: Air voids; Reinstatement; Testing; Density; Performance testing, SROH

1. Introduction

In 2013, a consortium of 23 members, representing utility undertakers, contractors, bituminous material suppliers, and a compaction equipment supplier instigated a university lead research project on utility reinstatement. Membership of the consortium includes representation from the gas, water, electric and telecommunications sectors in the UK.

Coring (the taking of samples of asphalt materials) programmes of utility reinstatements initiated by Local Authorities have been identifying consistent failure in respect of air voids (AV) content in surface course material of footways when assessed against the requirement of the Specification for the Reinstatement of Openings in Highways (SROH) (Department for Transport, 2010) for air voids content compliance only. This is an issue which currently affects all National Joint Utility Group (NJUG) members, presenting a significant and growing challenge as more Local Authorities apply the SROH air voids content standard to utility reinstatements.

Initial research focussed on critical analysis of the available published Standards and previous reinstatements trial results in the UK and the findings resulted in the publication of a white paper in 2014. The key features of the white paper was accepted as a technical article in a peer reviewed international journal (Sadique et al., 2015). Among various findings, two significant outcomes of this initial research were;

I. The use of air voids content determination on single cores is so inaccurate as to make compliance largely a matter of chance, as a result of compounding errors in the measurement of bulk density and maximum density. The use of air voids content other than for design mixtures, does not comply with UK best practice as outlined in BS594987: 2010, due to the within mix variability for recipe mixtures and the use of hand laying as the principal method of installation. The use of a measured in-situ air voids content criteria in a Specification for footway reinstatements cannot be sustained on technical grounds;

II. The linkage between AV content and durability in recipe mixed hand laid reinstatements with the limits currently in SROH is non–proven.

2. Research objective and method

To understand if the current AV compliance criterion in the SROH is able to provide a reliable indication of structural resilience throughout its service period, the following objectives were identified for further investigation within this research:

- To identify if AV content varied significantly in a small reinstatement, pairs of cores were collected by two different independent laboratories from different reported failed (in terms of AV compliance) reinstatement sites;
- To determine the susceptibility of the reinstatement to deform under load, samples were collected from previously failed (in terms of AV compliance) reinstatement sites for testing under wheel tracking ;
- To collect and review information from utility undertakers and contractors relating to in-situ performance of the reinstatement that previously reported failed (in terms of AV compliance) by the Highways Authority.

The comparison sites for coring were selected in such a way that a range of road categories as well as differentials in degrees of failure (in terms of AV content only) from minor to extensive were examined. The comparison cores were all taken in close proximity (within 100 mm) to ensure that both cored test sites had been similarly compacted with similar (almost identical) material.

Wheel tracking test was scheduled to provide reasonable measure of the future performance of good in-service performing footway reinstatements despite reported high AV. he specimen extraction, preparation and testing for permanent deformation was carried out using wheel tracking apparatus in accordance with the procedure stated in the BS EN 12697-22 (European Committee For Standardization, 2003b) using small size device.

In-situ performance of a number of reinstatements (footway and carriageway) of varying ages and varying only non-compliant AV contents were visually inspected by the respective undertakers/contractors across five various parts of the country (as shown in Figure 8) and reports were collated.

3. Uncertainty and reliability of AV content testing in hand laid reinstatement

The general approach to evaluating and expressing uncertainty in testing outlined in UKAS publication reference LAB-12 (United Kingdom Accreditation Service, 2000) was based on the recommendation produced in the guide by the International Bureau of Weights and Measures (BIPM et al., 2008). Providing a measure of uncertainty that defines an interval about the measurement result that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand was stated in this guide. Moreover, the general requirement for the estimation and reporting of uncertainty of measurement by all accredited laboratories has been specified with the implementation of the International Standard ISO/IEC 17025 (Birch, 2003), encompassing a number of influence quantities that affect the result obtained for the measurand in the case of uncertainty evaluation process.

To quantify the agreement and reliability of measurements made by any particular method or observer/s, a repeatability and reproducibility study of that measurement should be investigated (Bartlett and Frost, 2008). The repeatability and reproducibility interval for testing air voids content has been specified in the Standard BS EN 12697-8 (European Committee For Standardization, 2003a) by multiplying the respective standard deviation with 2.77. It is similar to the statistical estimate of a 95% confidence interval for the difference between two readings stated by ASTM Standard (Ullman, 2009). Based on this, the reproducibility statement for single coring results on identical test material reported by two laboratories, the air void contents should differ by no more than 2.2% on average on 95% of occasions (British Standards Institution, 1987, Bartlett and Frost, 2008).

To investigate reproducibility, five sites (C1 to C5) were selected from an undertaker's reinstatement where cores were taken by three UKAS accredited laboratories. The locations of the cores have been shown in Figure 1 to Figure 3. In order to keep the name of the laboratory performing the testing anonymous, the three test houses were named as Lab A, Lab B and Lab C. During this test, the maximum density was determined in accordance with EN 12697-5 (procedure A) and the bulk density was determined in accordance with EN12697-6 (procedure C) in all laboratories. The details of the test results have been shown in Table 1.

Figure 1: Location of three cores taken by Lab A, Lab B and Lab C from site C1 and C2

Figure 2: Location of three cores taken by Lab A, Lab B and Lab C from site C3 and C5

Figure 3: Location of three cores taken by Lab A, Lab B and Lab C from site C4

A strong statistical tool “paired t-test” was used to compare the AV content of two adjacent cores measured by two laboratories. During the t-test, the null hypothesis was assumed that the mean of two paired samples are equal and the alternative hypothesis was assumed that the means of two paired samples are not equal. The appropriate hypothesis was tested in the form of a probability - the p-value (significance 2-tailed) at 5% level of significance. If p is small ($p < 0.05$), the findings are unlikely to have arisen by chance and there is moderate evidence against the null hypothesis in favour of the alternative. If p is large ($p > 0.05$), the observed difference is plausibly a chance finding and there is no evidence against the null hypothesis. Smaller p-values ($p <0.01$) are sometimes called ‘highly significant’ because they indicate that the observed difference would happen less than once in a hundred times if there was really no true difference.

Table 1: The core test results obtained from five sites from an undertaker's reinstatements

Site Ref	Reinstatement Materials	Lab A				Lab B				Lab C			
		Depth (mm)	Voids (%)	Max (Mg/m^3)	Bulk (Mg/m^3)	Depth (mm)	**Voids (%)**	Max (Mg/m^3)	Bulk (Mg/m^3)	Depth (mm)	Voids (%)	Max (Mg/m^3)	Bulk (Mg/m^3)
C1	AC6 DSC	**69.0**	**20.7**	2.554	2.027	**113.0**	**12.2**	2.467	2.165	**110.0**	**15.4**	2.472	2.092
C2	AC6 DSC	**90.0**	**19.2**	2.564	2.074	**95.0**	**13.2**	2.486	2.157	**87.5**	**9.0**	2.477	2.255
C3	AC6 DSC	**92.0**	**21.8**	2.496	1.953	**89.0**	**6.3**	2.39	2.239	**90.0**	**10.7**	2.45	2.188
C4	AC6 DSC	**53.0**	**14.7**	2.497	2.132	**61.0**	**7.0**	2.361	2.195	**60.0**	**10.1**	2.406	2.163
C5	AC10 DSC	**102.0**	**15.0**	2.561	2.179	**80.0**	**8.0**	2.507	2.306	**50.0**	**8.0**	2.451	2.254
	AC20 DBC	**49.0**	**8.7**	2.6	2.375	**70.0**	**5.5**	2.543	2.402	**105.0**	**7.0**	2.5	2.325

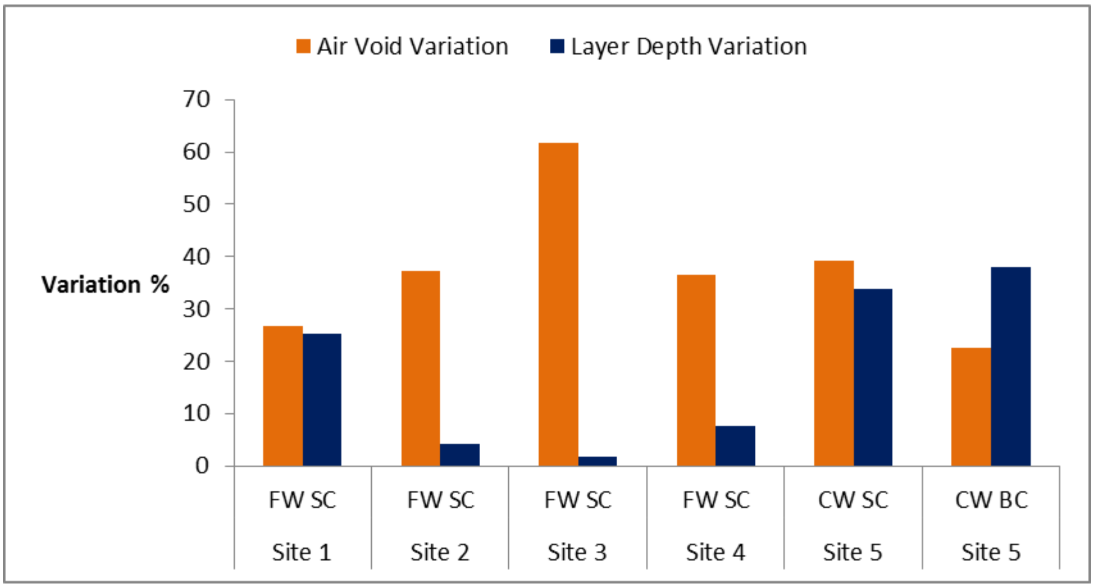

Figure 4: The relation between AV content and layer depth variation

The wide variation of results relating to bulk and maximum densities as well as layer depths obtained from three cores (located approximately 100 mm apart) are evident from Table 1, although the three test houses followed the same Standards and procedure stated in SROH. Although, the total recorded depths by the three test houses for site C5 were almost identical, however, in the case of identifying surface and binder coarse materials from same core), a wide variance of layer depth between three test houses were revealed. The maximum variations of layer depths were recorded in the case of laboratory A and laboratory C, where surface and binder coarse were varied by 104% and 114% respectively. As the AV content compliant requirement in SROH for binder and surface coarse materials are not similar, hence this high range of observational variation will essentially affect the assessment outcome of a reinstatement. Moreover, no recurring correlation was exists between the variation of layer depth and corresponding air void content in this comparison core analysis as shown in Figure 4.

The paired sample test results of Lab A-B, Lab B-C and Lab C-A have been shown in Table 2-4 respectively. It is evident from the t-test results that, statistically significant, $p < 0.05$ ($p = 0.005$ and 0.006, $t = 4.75$ and 4.58) differences of measured air void were revealed in the case of core results of Lab A-B and Lab A-C when compared. Furthermore, the 95% confidence interval of the difference lies in the range of 3.6% to 12.3% and 2.9% to 10.3% respectively (as shown in Table 2 and 3). However, non-significant, $p > 0.05$ ($p = 0.342$, $t = 1.05$) difference together with a lower range of the 95% confidence interval of the difference was reported in the case of Lab B-C (Table 4). The evidence of no recurring correlation between the variation of layer depth and corresponding air void content intensifies the significance of the reliable difference values that were observed in the t-test analysis for core results of Lab A-B and Lab A-C.

Table 2: Paired t-test for Laboratory A and B

	Paired Differences							
	Mean	Standard Deviation	Standard Error Mean	95% Confidence Interval of the Differences				
				Lower	Upper	t	df	**Significance (2-tailed)**
Pair 1 **Lab A – Lab B**	7.98	4.11	1.68	3.66	12.30	**4.75**	5	**0.005**

Table 3: Paired t-test for Laboratory A and C

	Paired Differences					t	df	**Significance (2-tailed)**
	Mean	Standard Deviation	Standard Error Mean	95% Confidence Interval of the Differences				
				Lower	Upper			
Pair 1 **Lab A – Lab C**	6.65	3.55	1.45	2.92	10.37	**4.58**	5	**0.006**

Table 4: Paired t-test for Laboratory B and C

	Paired Differences					t	df	**Significance (2-tailed)**
	Mean	Standard Deviation	Standard Error Mean	95% Confidence Interval of the Differences				
				Lower	Upper			
Pair 1 **Lab B – Lab C**	-1.33	3.11	1.27	-4.56	1.93	**-1.05**	5	**0.342**

In order to make the analysis more assured and representative, further a total of 68 pairs of comparison cores were taken (including the above 5 sites) from the reinstatements constructed by different undertakers within different parts of the country following the same procedure stated above. In this case, comparisons were made between the cores taken by Lab A (same as above) and those taken by different laboratories (here termed as Lab X). Detailed results from the 68 pairs of cores have been tabulated in Appendix A. The distribution of differences of AV content between the two laboratories in 68 reinstatement sites was found to be approximately normal as shown in Figure 5.

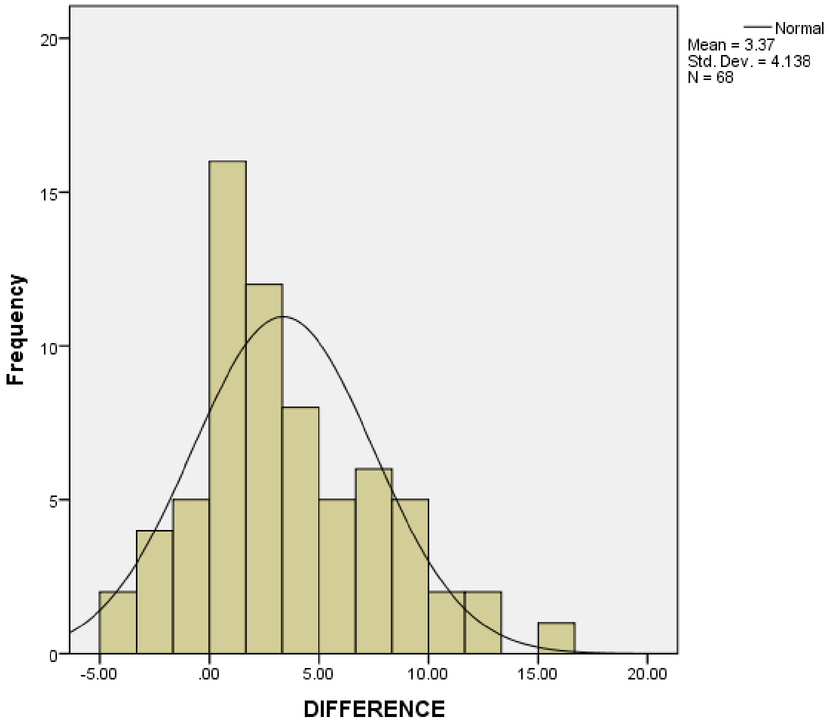

Figure 5: Histogram of differences in AV content between Lab A and Lab X

The repeatability and reproducibility interval for testing AV content has been specified in the Standard BS EN 12697-8 (European Committee For Standardization, 2003a) is similar to the statistical estimate of a 95% confidence interval for the difference between two readings stated by ASTM Standard (Ullman, 2009). Based on this, the reproducibility statement for single coring result on identical test material reported by two laboratories should differ by no more than 2.2% on average on 95% of occasions (British Standards Institution, 1987, Bartlett and Frost, 2008). However, in practice, acceptance (pass or fail against SROH AV content requirement) are evaluated based on the result from single core.

Table 5: Paired t-test for Laboratory A and X from 68 reinstatement sites

	Paired Differences							
	Mean	Standard Deviation	Standard Error Mean	95% Confidence Interval of the Differences				
				Lower	Upper	t	df	**Significance (2-tailed)**
Pair 1 **Lab A – Lab X**	3.36	4.13	0.501	2.36	4.37	**6.71**	67	**0.000**

The data from a t-test analysis of the all 68 pairs of cores has been recorded in Table 5. A closer examination to the "paired sample T-Test" among the pairs taken from 68 different sites reveals that, not only statistically significant (T = 6.7 and p = 0.000) difference between the AV content measured by two laboratories exists, but also the range of difference at 95% confidence level lies between 2.3% to 4.3%.This wide range exceeds the 2.2% reproducibility limit set by the British Standard. Moreover, acknowledging the proficiency relating to coring and testing procedures of the UKAS accredited test houses in this research, the extremely low intra-class correlation coefficient from reliability analysis (as shown in Table 6) of 68 pairs of air void content results inevitably indicates the poor reliability of the coring method.

Table 6: Intra-class correlation coefficient (ICC) analysis of 68 pairs of AV content for reliability test

		95% Confidence Interval		F Test with True Value 0			
	Intra-class Correlation	Lower Bound	Upper Bound	Value	df1	df2	**Significance**
Single Measures	**0.399**	0.035	0.640	3.190	67	67	**0.000**
Average Measures	**0.571**	0.068	0.780	3.190	67	67	**0.000**

In order to assess agreement between the measurements and presence of any bias within the reported AV content results produced by two test houses, a Bland-Altman plot was conducted. The Bland-Altman plot (Altman and Bland, 1983) and analysis is used to compare two measurements of the same variable and is a commonly referred method of comparison technique (Bartlett and Frost, 2008). The Bland-Altman plot of the AV content results from two independent laboratories has been shown in Figure 6. The solid green line indicates the mean of the paired differences (Lab A – Lab X) of air void content (3.36%) and its distance from zero provides the amount of bias between the two laboratories. The variability of the differences between the results of two laboratories indicates how well the method of assessment by AV content agrees. The limits of agreement give a range within which 95%

of future differences in measurements between two core results by two different laboratories would be expected to lie. The limits of agreement in this study were found to be in the range of -4.73% to 11.45% (mean difference ± 1.96x SD of differences). So, AV content measured by laboratory A may be 4.73% below or 11.45% above laboratory X on 95% of occasions in future (represented by dashed lines in Figure 6).

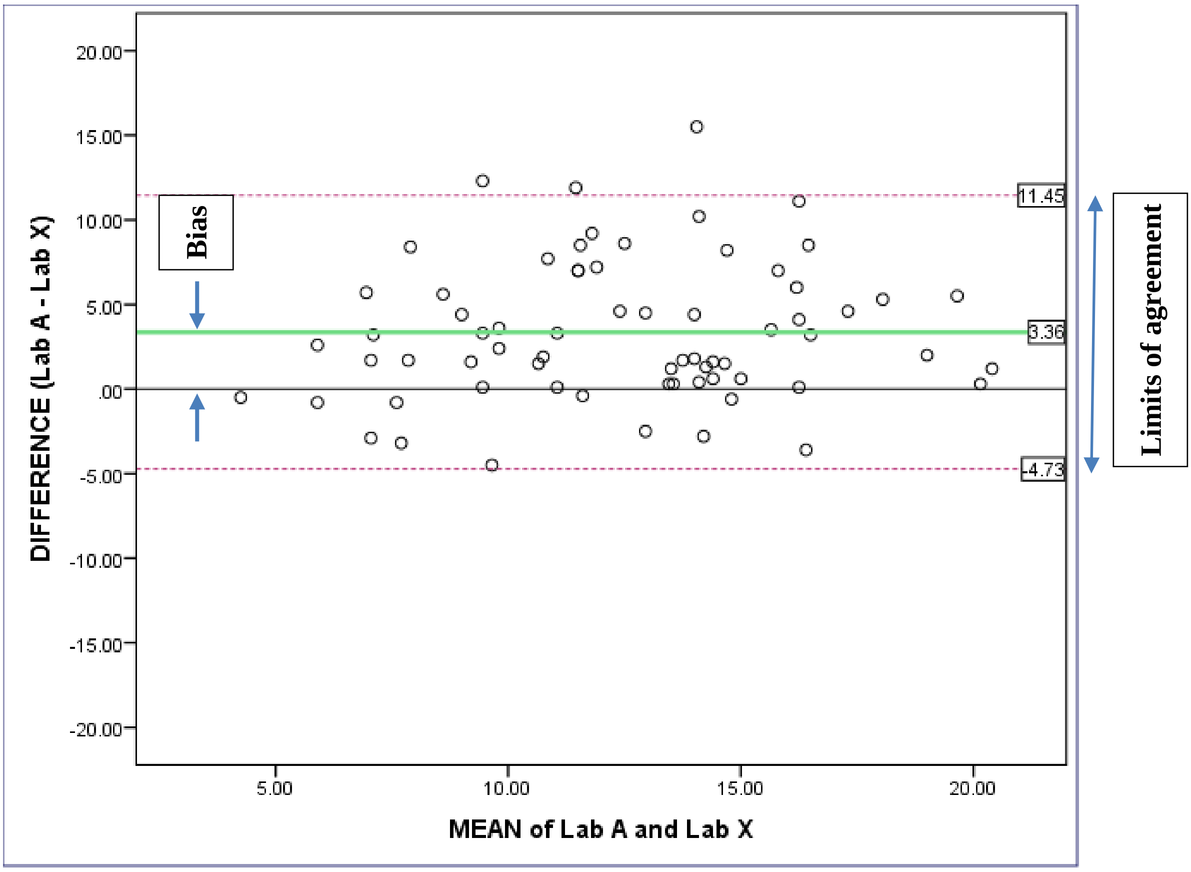

Figure 6: Differences in AV content measured by Lab A and Lab X against their means (Bland-Altman plot)

Density plays a very important role in AV content calculation and a small change in bulk and/or maximum density values affects the AV content significantly. Variations in the maximum and bulk densities between two specimens, which may be reasonably considered as the same (100 mm apart in this case) should rarely if ever occur and if they do they should be such a minimal amount that they do not impact on the integrity of the test results. However, it was revealed in this research that, in the case of algebraic difference of measured densities between two test houses, the Lab A measured maximum density values were higher than Lab X for the same parameter in 88% of cases, whereas in 72% of cases, Lab A measured bulk density values were lower than Lab X (as shown in Table 7). It was also observed that the differences (Lab A-Lab X) of densities were not compensating each other when total 68 sites were considered. As a consequence of this compounding difference of densities, a statistically significant difference between the AV content measured by two laboratories was revealed in a t-test analysis and the amount of bias in Bland-Altman analysis was in accordance with this finding. However, as both test houses were UKAS accredited and followed the same procedures stated in the relevant British standards and SROH, the absolute difference between each pair (locating 100 mm apart) were measured as shown in Table 7. Furthermore, applying "Microsoft Excel Data Solver" tool was employed to investigate the sensitivity of the results obtained from 68 sites summarised in Table 7. Based on the reported pattern of differences of maximum and bulk densities, 0.066 Mg/m^3 (as close as possible to 0.0644 Mg/m^3) decrease of maximum density and 0.057 Mg/m^3 (as close as possible to 0.0634 Mg/m^3) increase of bulk density was used during the AV content sensitivity (nearest to one decimal place) test and the following sensitivity were reported:

- Only 0.100 Mg/m^3 decrease of maximum density decreases AV by 3.6%
- Only 0.100 Mg/m^3 increase of bulk density decreases AV by 4.0%
- Combined, 0.066 Mg/m^3 decrease of maximum density and 0.057 Mg/m^3 increase of bulk density , decreases AV by 4.7%

Table 7: Summary of density measurements from 68 sites measured by Lab A and Lab X.

		Lab A – Lab X	
Average algebraic difference	Maximum Density	0.0576 Mg/m^3	In 88% cases, Lab A measured maximum density values were **higher** than Lab B
Average algebraic difference	Bulk Density	-0.0321 Mg/m^3	In 72% cases, Lab A measured bulk density values were **lower** than Lab B
Average absolute difference	Maximum Density	0.0644 Mg/m^3	
Average absolute difference	Bulk Density	0.0634 Mg/m^3	

Through implementing the 2nd Edition of the SROH, the Department for Transport introduced an end result specification (ERS) in place of method specification for assessing asphalt material. However, the compaction for asphalt material for major road construction in British Standards BS 594987 (European Committee For Standardization, 2010) is assessed by stating the following:

> *"End result compaction shall be applied to designed dense base and binder AC mixtures which have been type tested in accordance with BS EN 13108-20. A method of compaction shall be adopted and detailed in a suitable quality plan so as to ensure that the void content of the finished mat conforms to the required limits on void content.*
>
> *NOTE: This method is applicable for works intended to carry heavy traffic. The scale of works should be such as to justify the cost of testing and control (clause 9.5.1.1).*"

Though, the following note has been quoted concerning the compaction of asphalt materials in BS 594987:

> *“End result compaction is more appropriate for machine‑laid work on major road contracts”* (clause 9.1).

According to SROH A2.0, the reinstatement work in footways and carriageways is considered to be carried out in confined or restricted areas. Hence, it is highly anticipated that, due to the method and nature of utility reinstatement construction (transportation from plant, unloading, laying and compaction in restricted areas), the homogeneity of the asphalt mixtures is likely to be distorted and as a consequence resulting maximum density will be varied within the reinstatement. The wide-ranging maximum density (average difference = 0.0481 Mg/m^3) reported in every instance in the above comparison pair coring experiments also meaningfully rationalizes the distorted homogeneity of materials. Material homogeneity was also specified as main criteria to consider a spot sample as average sample in relevant British Standards (European Committee For Standardization, 2001). Although not only maximum density but also bulk density of adjacent cores located only 100 mm apart were found to be varied in the case of every pair (average difference = 0.0474 Mg/m^3) in this study. So it can be stated that within a pair of adjacent cores, the variation of maximum density originates from the distorted material homogeneity whereas, due to intrinsic biasness within the bulk density testing procedure in the relevant Standard, the bulk density diverges from each other.

The lack of material homogeneity is very unlikely in the case of any machine‑laid asphalt work on major construction. Therefore, appropriateness relating to the compliance assessed only by measuring in-situ AV content using the material and method of construction quoted in the SROH is not justified. Moreover, the use of air voids content requirement and associated testing regime for recipe mixed hand laid reinstatement works is acknowledged to be not totally suitable in the relevant British Standard due to service load (footways), scale of work (utility reinstatement), nature of construction (hand laid) and material used (recipe mixed).

Hence, at 95% level of significance, there exists enough evidence to conclude that, due to high uncertainty, very low repeatability and reproducibility and poor reliability with high chances of bias, the assessment of hand laid reinstatement work by AV testing will expose both the contractor and the client to unacceptable risk.

4. Durability of non-compliant footways reinstatement

For assessing the durability of adhesion in high modulus base and binder course mixture, the saturation ageing tensile stiffness (SATS) test was developed for trunk roads and motorways in the UK. However, the applicability of this test method was limited to bituminous specimens with consistent air void contents and hard binder, air void contents between 6 % and 10% and 10/20 pen hard paving grade bitumen (European Committee For Standardization, 2012). Also this test generally involves specimens cored from a slab manufactured using a laboratory roller compactor as research data on the performance of in-service specimens in the SATS test is currently unavailable.

The wheel tracking test is a widely used performance related test, which is known to correlate with an engineering property to predict performance and durability. The susceptibility of bituminous materials to deform is assessed by the wheel tracking test at constant temperature. The limiting value for resistance to permanent deformation has been specified in PD 6691 (European Committee for Standardization, 2015), appropriate only for the carriageway designed for class 1 (1001 to 2000 commercial vehicle/lane/day) and class 2 (2001 to 4000 commercial vehicle/lane/day) traffic sites as classified in MCHW series 900 (Highways Agency, 2008).

Instead, the lowest penetration grade bitumen permitted in SROH is 40/60 and this research was intended to investigate both the performance and the structural integrity of in-service (not laboratory prepared) footway reinstatement (generally not high modulus) containing high air void (more than 13%). Also, no performance or durability related test method for base and binder course materials has been specified in SROH except measuring the resistance to permanent deformation of surface course mixture. Moreover, there are instances of mixtures performing poorly in the SATS test but demonstrating a proven record of good performance in-service (and vice versa) have also been reported (Nicholls et al., 2011). Hence to check the resilience of in-service low modulus asphalt materials (generally surface course using 100/150 binder with non-consistent AV as result of recipe mixed), but already experienced sufficient (in the range of 6 years to 1.5 years) real life ageing, weathering and environmental loading; a more reliable and realistic judgement cannot be made from SATS test. Therefore, a wheel tracking test was scheduled to provide reasonable measure of the future performance of good in-service performing footway reinstatements despite reported high AV. Although, requirement for assessing the resistance to permanent deformation in non-traffic sites has not been stated in relevant British Standards.

The specimen extraction, preparation and testing for permanent deformation was carried out using wheel tracking apparatus in accordance with the procedure stated in the BS EN 12697-22 (European Committee For Standardization, 2003b) using small size device. The wheel tracking slope (WTS) and proportional rut depth (PRD) were measured using procedure B in air, whereas the wheel tracking rate (WTR) was measured according to procedure A. The temperature for testing as well as conditioning up to required duration was kept at 45°C as stated in PD 6691 for moderate to heavily stressed sites (class 1). Specimens were cut of sufficient size from the reinstatement to enable them to be sawn to form a rectangular test specimen of 260 mm × 300 mm for small size devices. Different stages of the sample extraction and testing have been shown in Figure 7. The results for one sample from each site have been tabulated in Table 8. Compliance relating to wheel tracking slope (WTS) was reported in all cases, and reasonable compliance relating to proportional rut depth (PRD) was reported in two cases despite the reinstatement were having 20.1% and 17.7% AV content (reported by a UKAS accredited test house) and good in-service performance against 6 years and 1.5 years of ageing, weathering, oxidation and wearing.

Withstanding occasional overrun by non-commercial vehicles (less than 1.5 tonnes unladen) has been specified as performance compliance for any footway reinstatement in SROH (S2.1.6). However, complying the requirement relating with WTS by non-compliant reinstatement (in terms of only AV content) in this research predictably indicate either:

- The reported AV content is extremely over estimated;

  or

- The linkage between the reinstatement with non-compliant AV and its impact on footways durability is non-proven

Figure 7: Specimen extraction and testing in wheel tracking apparatus

Table 8: Wheel tracking test results on footways specimens with non-compliant AV content

Sample reference	Reinstatement date	% AV reported by HA	Coring date	WTR test date	WTR results	Maximum allowable limit in PD6691
WTR-C2	February ‘14	19.3%	September ‘14	March ‘15	WTS_{Air} = 0.729 mm/1000 cycle PRD_{Air} = 15.68% WTR_{Air} = 1.60 μm/cycle	WTS_{Air} = 1.0 mm/1000 cycle PRD_{Air}= 9.0%
WTR-3rd - 2	April ‘2009	20.1%	July ‘14	March ‘15	WTS_{Air} = 0.414 mm/1000 cycle PRD_{Air} = 10.42% WTR_{Air} = 0.53 μm/cycle	
WTR-A2	November ‘13	17.7%	November ‘14	April ‘15	WTS_{Air} = 0.184 mm/1000 cycle PRD_{Air} = 4.35% WTR_{Air} = 0.747 μm/cycle	

5. In-situ performance of reinstatement reported non-compliant AV

In-situ performance of a number of reinstatements (footway and carriageway) of varying ages and varying only non-compliant AV contents were visually inspected by the respective undertakers/contractors across five various parts of the country (as shown in Figure 8) and reports were collated. The samples were selected at random and include reinstatements with AV in the range of 14.4% to 25.9% and in-situ performance life was in the range of 1.5 years to 10 years. During this range of assessment period, the UK experienced various extreme weather events including record rainfall, flood, wettest winter, record low temperatures, exceptionally heavy snow fall and warmest month on record (Met Office, 2015).

Evidence was collated from approaching 50 sites across various areas of the country and no visual failures were recorded that would have breached the performance tolerance permitted by section S2 of SROH. Moreover, in many instances the reinstated area was performing better than the surrounding highway and none of the reinstatements were found to be inferior in any respect to the condition of the adjacent surface. Only three typical visual in-situ assessments have been shown in Figure 9 and the location of the reports are available from all assessments, as are the associated UKAS air void testing certificates (from the original core tests). Hence the resilience shown during this in-situ performance assessment by these non-compliant reinstatements across the country in service performance against ageing, weathering, oxidation, wear and different extreme environmental loading was also in conformity with the findings reported by the wheel tracking test observation in this research.

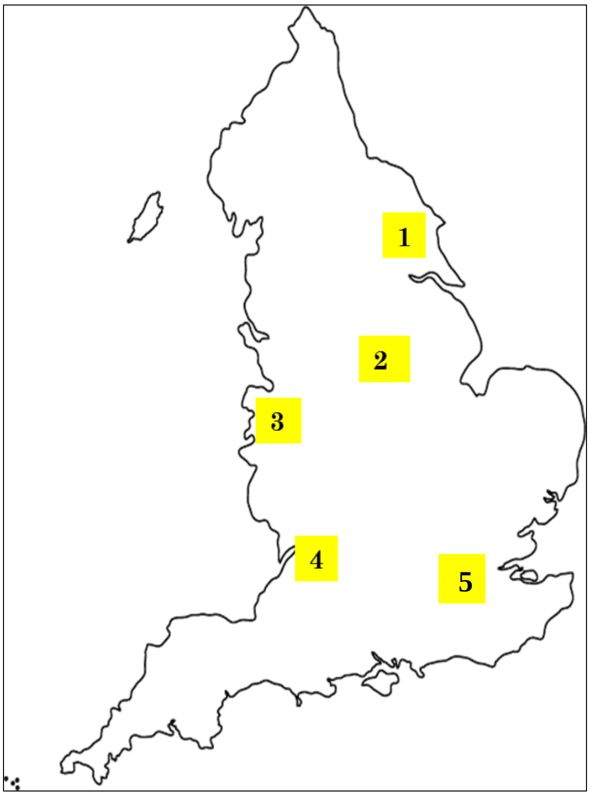

Figure 8: Location of visually assessed in-situ performance of reinstatement across the UK

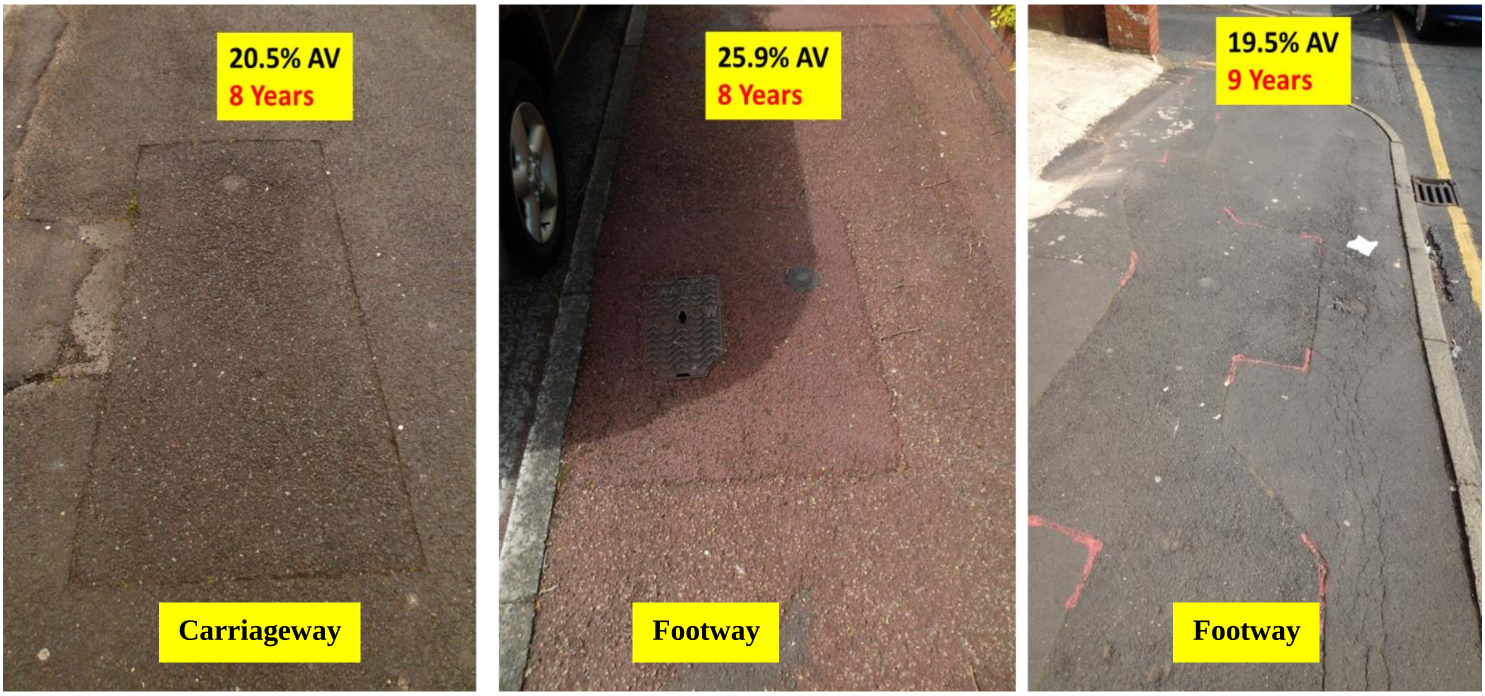

Figure 9: Visual in-situ performance of different reinstatement containing non-compliant AV

6. UKAS Position Statement on Reliability of using AV Content for Assessing Utility Reinstatement

Based on the above findings, this research forwarded a letter to the Technical Advisory Committee for Construction Industry in UKAS requesting their thoughts on the issues outlined above with focus on the inconstancies highlighted within UKAS accredited providers. Accordingly the research team shared the research findings as stated in this paper above. After the discussion on the above findings from this research, the UKAS Technical Advisory provided following position statement to the research.

"A representative core sample taken and subjected to testing by a UKAS an accredited laboratory in accordance with BS EN 12697 for hand laid recipe mixtures may only provide confidence in the sample tested meeting the requirements of the Specification for Reinstatement of Openings in Highways, and may therefore not be considered for the integrity in conformity of the whole reinstatement. In contrast, machine laid work is generally homogeneous and so the analysis of a

single core is may provide a result that is representative of the material in the whole reinstatement than would be the case for hand laid material. However, whatever the method of laying, test results can only accurately represent the sample that has been analysed and cannot validly be used to represent the composition of adjacent material" (Giles and Chapman, 2016).

7. Conclusion

The comparison pairs of cores (each 100mm apart) from 68 reinstatements sites from various parts of the UK revealed that, the compounding consequences of generic non-homogeneous characteristics of hand laid recipe mixed materials (specified in SROH) and high likeliness of being biased during AV testing makes the coring method extremely unreliable with very low repeatability and reproducibility. The position statement provided by UKAS technical committee (as reported in the previous section) also absolutely in accordance with the above findings (distorted homogeneity of hand laid recipe mixed material).

Furthermore, the in-situ performance shown by from 50 reinstatements after experiencing 1.5 years to 10 years real life aging from various parts of the UK predictably indicates that either the linkage between the reinstatement with non-compliant AV and its impact on footways durability is non-proven or the reported AV content is extremely over estimated. The resilience shown by the non-compliant AV content reinstatement to withstand structural loading as well as extreme environmental loading beyond the guarantee period across the country validates the above finding.

It is envisaged that numerous reinstatements will have to be reworked based on an assessment method which is itself not only unreliable but also suffering from non-compliant precision relating to the British Standards. The revealed inherent embedded biasness as well as unreliability of current assessment method of SROH, for a hand laid bituminous work where non-homogenous materials are likely, situating both the contractor and the client at unacceptable risk and costing utilities, contractors and the community without any additional benefit in performance.

A specification should be realistic, practical, and sustainable and be able to predict performance. The current specification for a hand laid recipe mixed material based upon coring for air void content, fails on all of these basic requirements. It could lead to a very wide range of unpredicted outcomes, putting both the contractor and the client at unacceptable risk.

8. Acknowledgements

The research team wish to thank the National Joint Utilities Group (NJUG) and consortium members who provided invaluable support to this research project during the study. Their support is thankfully acknowledged.

References

ALTMAN, D. G. & BLAND, J. M. 1983. Measurement in medicine: the analysis of method comparison studies. *The statistician*, **,** 307-317.

BARTLETT, J. & FROST, C. 2008. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound in Obstetrics & Gynecology,* 31**,** 466-475.

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP & OIML 2008. Evaluation of measurement Data-Guide to the expression of uncertainty in measurement-JCGM 100:2008. Joint Committee for Guides in Metrology.

BIRCH, K. 2003. Estimating Uncertainties in Testing - An Intermediate Guide to Estimating and Reporting Uncertainty of Measurement in Testing, Measurement Good Practice Guide No. 36. Middlesex: British Measurement and Testing Association.

BRITISH STANDARDS INSTITUTION 1987. Precision of test methods-Part 1: Guide for the determination of repeatability and reproducibility for a standard test method by inter-laboratory tests - BS 5497-1: 1987.

DEPARTMENT FOR TRANSPORT 2010. Specification for the Reinstatement of Openings in Highways (SROH), 3rd Edition. The Stationery Office (TSO).

EUROPEAN COMMITTEE FOR STANDARDIZATION 2001. Bituminous mixtures -Test methods for hot mix asphalt, Part 27: Sampling, BS EN 12697-27:2001.

EUROPEAN COMMITTEE FOR STANDARDIZATION 2003a. Bituminous mixtures -Test methods for hot mix asphalt : Part 8: Determination of void characteristics of bituminous specimens, BS EN 12697-8:2003. Brussels: British Standards Institutions.

EUROPEAN COMMITTEE FOR STANDARDIZATION 2003b. Bituminous mixtures -Test methods for hot mix asphalt, Part 22: Wheel tracking, BS EN 12697-22:2003. London.

EUROPEAN COMMITTEE FOR STANDARDIZATION 2010. Asphalt for roads and other paved areas – Specification for transport, laying, compaction and type testing protocols, BS 594987:2010. London: British Standard Insititution.

EUROPEAN COMMITTEE FOR STANDARDIZATION 2012. Bituminous mixtures -Test methods for hot mix asphalt, Part 45: Saturation Ageing Tensile Stiffness (SATS) conditioning test -BS EN 12697-45:2012.

EUROPEAN COMMITTEE FOR STANDARDIZATION 2015. Guidance on the use of BS EN 13108 Bituminous mixtures – Material specifications, PD 6691:2015. British Standards Institutions,.

GILES, S. & CHAPMAN, P. 2016. United Kingdom Accreditation Service (UKAS) Statement Regarding Core Samples to BS EN 12697.

HIGHWAYS AGENCY 2008. Manual of Contract Documents for Highway Works - Volume 2, Notes for Guidance on The Specificat Ion For Highway Works, Series NG 900. London.

MET OFFICE. 2015. *Historical Weather* [Online]. Available: http://www.metoffice.gov.uk/learning/library/publications/historical-facts [Accessed 01/06/2015 2015].

NICHOLLS, J. C., PRIME, J., MEITEI, B. & LOWE, A. 2011. Review of Saturation Ageing Tensile Stiffness (SATS) test for use in Ireland- Published project Report - PPR547. Transport Research Laboratory.

SADIQUE, M., AL-NAGEIM, H. & STOPPS, K. 2015. To investigate the fundamental causes of utility air voids content failures in asphalt layers to achieve Specification for the Reinstatement of Openings in Highways (SROH) compliant performance. *Construction and Building Materials,* 93**,** 595-607.

ULLMAN, N. 2009. What Are Repeatability and Reproducibility? Part 2: The E11 Viewpoint. Magazines & Newsletters / ASTM Standardization News

UNITED KINGDOM ACCREDITATION SERVICE 2000. The Expression of Uncertainty in Testing [UKAS Publication ref: LAB 12]. Middlesex: UKAS.

Appendix A

Core test results from different sites

Site Ref	Material	**Laboratory A**				**Laboratory X**			
		Depth	**Voids**	Max	Bulk	Depth	**Voids**	Max	Bulk
1	AC6 DSC	79	**14.6**	2.488	2.125	79	**12.9**	2.477	2.157
2	AC6 DSC	87	**15.3**	2.485	2.105	83	**14.7**	2.477	2.113
3	AC6 DSC	81	**15.2**	2.522	2.138	77	**13.6**	2.47	2.134
4	AC6 DSC	77	**14.9**	2.732	2.324	82	**13.1**	2.731	2.372
5	AC6 DSC	71	**13.6**	2.437	2.104	87	**13.3**	2.48	2.149
6	AC6 DSC	92	**14.3**	2.517	2.158	90	**13.9**	2.489	2.143
7	AC6 DSC	65	**14.5**	2.477	2.117	66	**15.1**	2.469	2.097
8	AC6 DSC	76	**14.7**	2.483	2.117	87	**14.1**	2.483	2.134
9	AC6 DSC	77	**16.3**	2.458	2.058	86	**16.2**	2.419	2.026
10	AC6 DSC	78	**15.2**	2.511	2.128	77	**10.7**	2.498	2.231
11	AC6 DSC	86	**14.1**	2.481	2.131	92	**12.9**	2.494	2.172
12	AC6 DSC	88	**15.4**	2.472	2.092	77	**13.9**	2.477	2.132
13	AC6 DSC	67	**14.6**	2.498	2.133	74	**18.2**	2.48	2.028
14	AC6 DSC	71	**20**	2.512	2.015	83	**18**	2.485	2.038
14	AC6 DSC	73	**21**	2.464	1.949	78	**19.8**	2.5	2.005
16	AC6 DSC	84	**18.1**	2.535	2.077	94	**14.9**	2.486	2.116
17	AC6 DSC	80	**18.3**	2.541	2.076	90	**14.2**	2.492	2.139
18A	AC20BC	154	**16.4**	2.489	2.082	94	**7.2**	2.41	2.237
19A	AC20BC	130	**11.2**	2.476	2.199	127	**6.8**	2.383	2.235
20A	AC20BC	110	**15.8**	2.647	2.23	118	**7.3**	2.48	2.308
21A	AC20BC	136	**5.6**	2.441	2.303	134	**8.5**	2.423	2.217
22A	AC20BC	160	**11**	2.473	2.202	154	**8.6**	2.415	2.208
23A	AC20BC	149	**11.1**	2.465	2.192	146	**11**	2.439	2.171
24A	AC20BC	137	**11.4**	2.453	2.173	132	**11.8**	2.41	2.126
25A	AC20 BC	161	**11.6**	2.462	2.176	162	**8**	2.469	2.235
26A	AC10 CSC	154	**5.5**	2.481	2.346	94	**6.3**	2.415	2.262
27A	AC10 CSC	117	**11.4**	2.436	2.159	156	**9.9**	2.403	2.165
28A	AC10 CSC	130	**7.2**	2.51	2.33	127	**8**	2.456	2.271
29A	AC10CSC	137	**18.8**	2.502	2.033	137	**10.6**	2.496	2.242
30A	AC10CSC	110	**4**	2.452	2.355	118	**4.5**	2.429	2.334
31A	AC10CSC	136	**11.7**	2.498	2.205	134	**9.8**	2.437	2.21
32A	AC10CSC	160	**9.5**	2.442	2.21	154	**9.4**	2.439	2.21
33A	AC10CSC	149	**10**	2.467	2.229	146	**8.4**	2.4	2.198
34A	AC10CSC	137	**7.4**	2.453	2.271	132	**11.9**	2.424	2.136
35A	AC10CSC	161	**6.1**	2.474	2.322	162	**9.3**	2.393	2.172
36B	AC6 DSC	**111**	**20.3**	2.498	1.991	**134**	**20**	2.53	2.024
37B	AC6 DSC	**88**	**14.9**	2.412	2.052	**92**	**13.6**	2.38	2.056
38B	AC6 DSC	**116**	**16.8**	2.514	2.092	**83**	**8.2**	2.415	2.218
39B	AC6 DSC	**109**	**13.7**	2.469	2.131	**83**	**13.4**	2.455	2.127
40B	AC10CSC	**102**	**12.7**	2.528	2.207	**74**	**9.4**	2.468	2.236
41B	SMA SC	**49**	**12.8**	2.517	2.194	**50**	**15.6**	2.523	2.13
42B	SMA SC	**122**	**11.7**	2.486	2.194	**52**	**14.2**	2.574	2.208
43N	AC6 DSC	**69**	**20.7**	2.554	2.027	**113**	**12.2**	2.467	2.165

Appendix A

Core test results from different sites

Site Ref	Material	Laboratory A				Laboratory X			
		Depth	**Voids**	Max	Bulk	Depth	**Voids**	Max	Bulk
44N	AC6 DSC	**90**	**19.2**	2.564	2.074	**95**	**13.2**	2.486	2.157
45N	AC6 DSC	**92**	**21.8**	2.496	1.953	**89**	**6.3**	2.39	2.239
46N	AC6 DSC	**53**	**14.7**	2.497	2.132	**61**	**7**	2.361	2.195
47N	AC10 CSC	**102**	**15**	2.561	2.179	**80**	**8**	2.507	2.306
48N	AC BC	**49**	**8.7**	2.6	2.375	**70**	**5.5**	2.543	2.402
49N	AC6 DSC	**76**	**19.3**	2.511	2.208	**73**	**12.3**	2.423	2.125
50N	AC6 DSC	**67**	**19.6**	2.497	2.009	**70**	**15**	2.447	2.08
51N	AC10 CSC	**61**	**22.4**	2.601	2.019	**60**	**16.9**	2.472	2.053
52N	AC BC	**157**	**7.9**	2.605	2.4	**180**	**6.2**	2.601	2.44
53N	AC6 DSC	**72**	**17.4**	2.573	2.127	**70**	**13.9**	2.511	2.162
54N	AC BC	**49**	**11.4**	2.626	2.329	**80**	**5.8**	2.534	2.387
55N	AC6 DSC	**69**	**20.7**	2.554	2.027	**110**	**15.4**	2.472	2.092
56N	AC6 DSC	**90**	**19.2**	2.564	2.074	**87.5**	**9**	2.477	2.255
57N	AC6 DSC	**92**	**21.8**	2.496	1.953	**90**	**10.7**	2.45	2.188
58N	AC6 DSC	**53**	**14.7**	2.497	2.132	**60**	**10.1**	2.406	2.163
59N	AC10 CSC	**102**	**15**	2.561	2.179	**50**	**8**	2.451	2.254
60N	AC BC	**49**	**8.7**	2.6	2.375	**105**	**7**	2.5	2.325
61E	HRA SC	**42.5**	**15.6**	2.415	2.039	**48**	**3.3**	2.127	2.056
62E	AC20 BC	**72**	**11.1**	2.591	2.305	**79**	**7.8**	2.516	2.32
63E	AC10 CSC	**41.25**	**15.5**	2.628	2.221	**44**	**8.3**	2.463	2.259
64E	AC20 BC	**88.25**	**7.2**	2.544	2.361	**73**	**4.6**	2.456	2.344
65E	AC20 BC	**62**	**17.4**	2.621	2.167	**67**	**5.5**	2.447	2.471
66E	AC20 BC	**54**	**16.2**	2.63	2.204	**42**	**11.8**	2.465	2.173
67E	HRA SC	**38.75**	**12.1**	2.533	2.227	**40**	**3.7**	2.367	2.28
68E	AC20 BC	**83**	**9.8**	2.576	2.325	**80**	**4.1**	2.418	2.319