

Uncertainty assessment of standard tests of bituminous mixtures

Marisol Barral¹, Carmen Calvo², Lucía Miranda³, Rafael Jiménez⁴, Francisco José Barceló³, Javier Loma⁵,
M^a Elena Hidalgo⁶, Santiago Tamayo⁷, Andrés Gómez⁸, José Berbis⁹, Carlos García¹⁰, Rafael Guillén¹¹,
Marisa Carmona¹², Antonio García¹³, Mar Subarroca¹⁴

¹ R&D, Campezo, San Sebastian, Spain

² Laboratory, Junta de Castilla y León, Valladolid, Spain

³ R&D, Repsol, Mostoles, Spain

⁴ Transport Laboratory, CEDEX, El Goloso, Spain

⁵ Laboratory, Asfalto y Construcciones Elsan, Madrid, Spain

⁶ Technical Department, Eiffage Infraestructuras, Madrid, Spain

⁷ Laboratory, Diputación de Alava, Vitoria, Spain

⁸ Laboratory, BECSA, Castellón, Spain

⁹ Technical Department, Pavasal, Valencia, Spain

¹⁰ Technical Department, Collosa, Valladolid, Spain

¹¹ Technical Department, Sacyr, Madrid, Spain

¹² Laboratory, Intromac, Cáceres, Spain

¹³ Laboratory, CEPSA, Alcalá de Henares, Spain

¹⁴ Technical Department, Sorigué, Lleida, Spain

Digital Object Identifier (DOI): dx.doi.org/10.14311/EE.2016.335

ABSTRACT

Characterization of any material requires the existence of test methods to quantify with the greatest reliability properties in the scope of the analysis. In the body of EN standards of asphalt mixtures, there is a section where the precision values shows repeatability and reproducibility of the test method. Test methods variability is usually based on interlaboratory exercises. However this section is not always completed, or the detailed data come from limited experiences or that do not correspond exactly to the method described in the standard. Information provided about the accuracy of the method should be considered very relevant.

This study is based on a work developed by Aleas (Group of Spanish Laboratories Associated to Asefma), a set of organizations specialized in the production and characterization of asphalt mixtures. Its aim is to evaluate the accuracy described in the test method "EN 12697-12 water sensitivity of bituminous mixtures", according to Method A, through a round robin test. In this EN standard, paragraph regarding accuracy indicates that the precision data have not been established, and there is a reference regarding similar indirect tensile tests.

In this paper it is estimated the precision of the test for two types of bituminous mixtures: BBTM type B and AC, through a round robin test which involved 15 laboratories. Final results are compared with those of the reference published in the EN standard. This study has also used to calculate an overall estimate of its uncertainty using a methodology developed within the working group ALEAS.

Keywords: Indirect tension, Performance testing, Quality assurance, Standardisation

1. INTRODUCTION

Whenever a result is obtained from a test, can be ensured that it is the real value of the measured characteristic? What degree of confidence does the collected value? These questions should be familiar for technicians who perform measurements or that receive a result from another laboratory.

Reliability of a measured value is very important for further analysis, especially when there are specifications that involve the validity of a specific value.

The result of a measure always has associated an uncertainty value that corresponds to the calculation of all possible errors that can be detected and quantified during the measurement. Therefore, the result of a measurement is an approximation to a real value that is completed with its uncertainty.

Surely any technician who handles test results knows about terms such as repeatability, reproducibility, accuracy, uncertainty, etc. All of them are associated with a result and its validity to ensure that in fact it as has been carried out providing all the guarantees and , therefore, showing its reliability.

These terms, which are not assessed in the majority of cases, are however fundamental to ensure that the provided value, for example, the amount of bitumen for bituminous mixture, or the value of sensitivity to water, are correct and true, especially when fulfilment of technical specifications is compulsory.

In addition, some of these terms, repeatability and reproducibility, appear, although not always, in the test standards specifying the requirement when a sample is tested several times in the same conditions or when it is tested by different laboratories. Although these terms are mentioned in the standards, it may be worthwhile asking whether repeatability and reproducibility are taken into account to accept the validity of a value using the corresponding test method. This question, which may be quite obvious, should be bore in mind by technicians who performs the test or that receives the final report, in order to ensure that a result is correct and therefore is valid.

Therefore, when a value from an analysis of a sample is obtained, can be fully ensured that a value of an essay is correct? Not always, until the entire test method is reviewed and the possible errors associated with the method have been evaluated, ending with the estimation of a value of "uncertainty" that always accompanies the final value.

2. WHAT IS MEASUREMENT UNCERTAINTY?

If we search for the definition of "measurement uncertainty" in the vocabulary of metrology international VIM [1], it indicates that it is a "non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, bases on the information ", where the term "measurand" indicates "quantity intended to be measured"

This concept of uncertainty indicates that any outcome is associated with a range that includes all possible sources of uncertainty that have been quantified for a method of a specific test, under specific conditions. The inclusion of that range in the final result provides an idea of the quality of the measurement.

And what are those contributions that are part of a given uncertainty value? These contributions can arise from:

- Feature of the magnitude that is measured, or the test method itself,
- Equipment used for measurement,
- Operator who performs the measurement,
- Environmental conditions required for the execution of the test.

Once potential sources of uncertainty are exposed, it is important to know their influence to quantify them and finally obtaining a value of uncertainty that is associated with the final result by completing the real value of a measurand.

Below a series of factors, which are the most common for the study of errors when making a measurement, are shown. However, it is important to note in the case of very specific methodologies, the existence of some other variable not considered within this list.

2.1 Measured magnitude.

Determining a magnitude or parameter may be more or less complicated according to the methodology used. An example may be the temperature measurement, a simple magnitude that only requires the use of a thermometer, inserting into the sample and acquiring the data. Nevertheless, other more complicated parameters, such as sensitivity to water of asphalt mixes, involve many steps until the final value is

expressed as a percentage of retained strength: manufacturing of test specimens, specimen's saturation, water conservation of the specimen before breaking and finally the tensile breaking specimens. In this example are many potential errors that can be made before achieving the final result compared to a single step method that only involves temperature measurement.

The more complex is the determination of a magnitude, the greater number of sources of uncertainty should be taken into account.

2.2 Equipment used for measurements.

The equipment used to carry out a measure can suffer disarrangements that could affect the quality of the result. For this reason they must be calibrated or verified at specified intervals, to ensure that the measurement is performed properly.

In the case of calibration operations, it is important to know its value of associated uncertainty. This value of uncertainty and the possible mismatch that can suffer equipment must be evaluated and quantified and taken as a variable to be considered in estimating the parameter to be measured.

2.3 Human errors

To make a measurement, there are available testing standards that are followed by technicians and the operations to be performed are described. It is therefore essential, especially in the case of more complex methodologies, a period of training and expertise to enable technicians knowing test procedures, and, later, reducing and limiting the influence of the human factor in committing errors during application of the test method. However, despite this, there is a human component in the performance of testing in terms of technical ability or skill to perform the measurement that leads to a potential error that must be quantified and will take part of the overall uncertainty.

How to quantify these "human error" is obtained from the realization of successive measurements under different conditions of repeatability and reproducibility, limiting its increasing the number of repetitions performed.

2.4 Environmental conditions

To carry out the measurement of a parameter is important to consider the most appropriate environmental conditions that are required depending on the sample and to determine the property. These conditions, despite being controlled during the process may suffer fluctuations and thus affect the final result of the measurement to be made. Therefore, it is important to quantify how variability of the environmental conditions affects the final result and taking it into account as a contribution to the variability in the final measurement.

Once the sources of uncertainty are defined, next question is: how these contributions are quantified?

In the literature there is a broad description of how the different types of errors are classified and how to calculate them. Summarizing they correspond to:

- Accidental or random errors that occur when successive measurements of the same parameter are carried out and where variability of different values, either above or below, is found. In this case there are not big differences from the central value and they occur randomly.
- Systematic errors, corresponding to a constant deviation in all actions always both above and below the actual value, being produced continuously and presenting a significant difference from the central value.

These errors, classified according to the deviation from the central value, they express in technical terms as "accuracy" and "precision" of a measurement. In Figure 1 the systematic errors and random errors associated with these expressions are shown.

In Figure 1, the circles represent the position of the value, being with the smaller circle the actual value. Crosses indicate repetitions of a measurement that may close, or not, to the central value, but whose values are inside one of the circles.

The number of circles represents the accepted range or scope of the measurement that has been estimated as acceptable. Outside these circles measurement values should not be accepted.

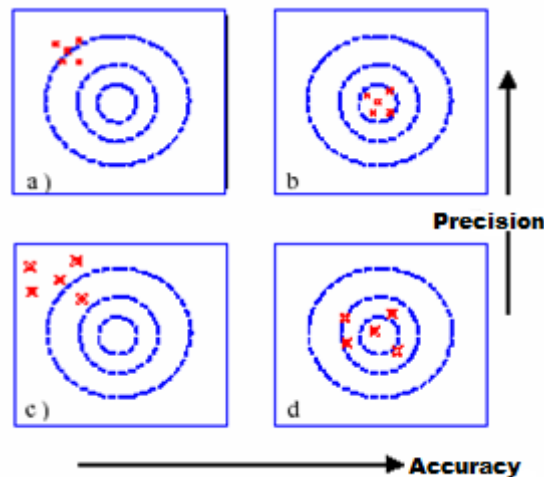


Figure 1. Graphic representation of “precision” and “accuracy”

Crosses disposal with respect to the central value, as well as their own dispersion within a certain circle, represent different types of errors.

Crosses that are more distanced from the central value corresponds to systematic errors and are always clearly different of the central value, such as circles depicted in a) and c). Furthermore, when the crosses are close to the central value but have small variations, above or below this value, correspond to random errors, as it is shown in b) and d).

Once classified the types of errors, it is obvious that they will smaller the lower dispersion is. Looking at Figure 1, two distinct types of dispersions are detected: one with respect to the central value and the other with respect to distance in repeated measurements.

Regarding the distance between the measurements, the closer are crosses the more precise is the measurement and it is associated with the term "precision". On the other hand, the more focused crosses are on actual value, the more accurate the measurement and it is associated with the term "accuracy".

Figure 1 shows the following situations: a) precise but inaccurate measurement, b) more accurate and with the same precision than a), c) less precise and less accurate, d) more accurate but less precise.

The ideal measurement would be one with 100% accuracy and 100% precision, and therefore there would be not any measurement error, a situation that is hard to find. So, it is necessary to calculate these errors and minimize them to finally get a value of uncertainty as low as possible and close to the ideal measurement.

Commonly the error term is usually associated to uncertainty, but it is important to differentiate these two. Error is the difference between a measured value and the actual value of a magnitude or property. If we know its value, we can use it as a correction of the result of a measurement. The value of uncertainty cannot be used to correct results. We can deal with different sources of uncertainty trying to minimize them, but they cannot be fully eliminated.

It is important to note that the value of the uncertainty of a result must always be less than the value of the magnitude being measured and should be suitable for the intended purpose thereof, allowing decision-making, for example, compliance with specifications.

3. HOW UNCERTAINTY CAN BE CALCULATED?

There is extensive literature regarding different methodologies to estimate the uncertainty of a measurement. All of them have an initial and essential step: to identify sources of uncertainty.

Having identified the different contributions to the uncertainty, the next step is to establish the method to quantify and to calculate a final value. To have a starting point in the evaluation and assessment of contributions, it is necessary having measurable and quantifiable information that can come from:

- Values of successive repetitions of a measurement that will reveal the dispersion of the results and thus the precision and accuracy compared to a central value, either under repeatability or reproducibility conditions.

- Information coming from certificates of calibration of equipment, taking into account the value of uncertainty that will differentiate between multiple equipment with the same characteristics which provides less uncertainty contribution.
- Information or data about the characteristics of the equipment: resolution, precision, accuracy bearing in mind the requirements described in the testing standards.

All this information, measurable and quantifiable, is used to apply a series of equations that usually are classified into two types, commonly referred to as "components":

- Components type A: They follow a normal distribution. They are calculated from the standard deviation and the number of repetitions that have been made for calculations. From this expression can be calculated uncertainties due to the repeatability and reproducibility.

$$U_{REP} = \frac{S_{CORR}}{\sqrt{n}}$$

- Components type B: They do not follow a normal distribution, but other (rectangular, triangular ...). In this expression, how to calculate the remaining uncertainties is shown as a range. For example, in the case of equipment that is calibrated and which has a value of uncertainty, the contribution is calculated using an expression such as:

$$U_c = \frac{I}{k}$$

Where:

- I: is the value of the uncertainty in the report of calibration
- k: is the degree of coverage for a given confidence level

Once sources of uncertainty have been quantified, according to their type and by using the corresponding expression, it requires using statistical techniques to combine all of them through a quadratic sum of the contributions. Thus a value of uncertainty, called the "combined uncertainty", is obtained.

A graphical representation of the value of the measurement and its combined uncertainty is shown in Figure 2, where the central value corresponds to the maximum value of the distribution and the uncertainty range would be on either side of this central value.

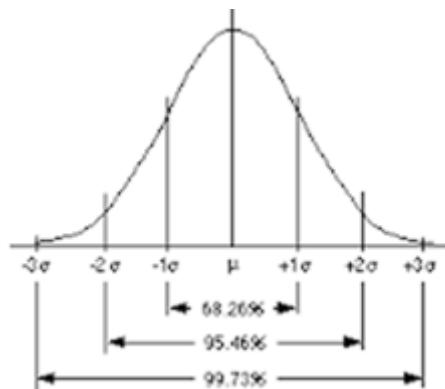


Figure 2. Distribution of uncertainty over the central value

Figure 2 shows three percentages corresponding to the degree of confidence where all possible measured values should be included and within which are considered acceptable values .

There are three classifications to define this interval, corresponding to:

- Taking an interval to a range of 68% of values that are accepted
- Taking an interval for 95% range of the values that are accepted
- Taking an interval for the range of 99% of the values that are accepted

Depending on the range of confidence to be chosen, the combined uncertainty is established multiplying its value by a factor corresponding to 1, 2 or 3, depending on whether the chosen interval is 68, 95 or 99%, getting the so called "expanded uncertainty". Usually, it is considered a confidence interval of 95%. Therefore, the expression of a value corresponds to an expression such as:

Value \pm uncertainty

It is important and necessary to emphasize that uncertainty is the uncertainty is typical of a laboratory and testing because of the uncertainty contributions are different by involving different equipment, different people, etc. However, to assess whether the value of uncertainty is consistent or not, there are tools such as round robin tests, where a laboratory can be compared against others, from the data provided on the analysis of the same sample, by using statistical techniques.. This comparison will allow a laboratory, which has an specific value of uncertainty, knowing if there is a lot of dispersion from the rest, proceed to assess its methodology and its "savoir faire" and check the possibility of reducing this uncertainty value to approach the other laboratories with which it has been compared.

4. WHAT INFORMATION IS PROVIDED BY MEASUREMENT UNCERTAINTY?

Knowing the value of uncertainty of a measurement is not easy to perform, as it is mentioned above, because in some cases it may be difficult to quantify the errors. For senders and receivers of technical information, knowing the value of uncertainty means that a comprehensive study of the entire analysis methodology has previously been made.

Besides, uncertainty value and its magnitude provide information on the characteristics of the test method for a particular sample type, and its ability to narrow or not the range of this value and, from here, to define consistent specifications.

If we take the whole process of estimation of uncertainty to the bituminous mixtures technology, it is important to consider several factors:

- Samples usually are heterogeneous so its reproducibility is difficult to be narrowed.
- The methodologies used in determining the properties are generally quite complex, with many stages, so the sources of uncertainty accumulate and become significant, leading to higher uncertainty values
- There are CE regulations, which includes limit values which are used to define the specifications of a material property

Taking into account these issues, information provided by measurement uncertainty is very important to define properly specifications. , If a value of uncertainty associated with a result exceeds the limit defined in these specifications, it shows that specifications been defined regardless of the overall result of a measurement (value + uncertainty), what would lead to inconsistency in the results.

5. METHODOLOGY FOR ESTIMATING THE UNCERTAINTY IN "DETERMINATION OF WATER SENSITIVITY IN A BITUMINOUS MIX" TEST

In order to show the importance of uncertainty in testing, and more specifically those used in asphalt mixtures, this paper show the main results of the round robin test about the determination of the water sensitivity (EN 12697-12) , carried out by 15 laboratories, belonging to Aleas¹ network, and compared with other international references. Additionally, the influence of the calculated levels of uncertainty when applied to the product standards of bituminous mixtures (EN 13108) and existing Spanish specifications will be studied.

¹ Aleas network is composed of 35 laboratories specialized in asphalt mixtures testing
<http://www.asefma.es/aleas/>

Within a working group of the ALEAS network a methodology [2] for the estimation of uncertainty the main tests of bituminous mixes has been developed. In this case we will focus in EN 12697-12.

This calculation is performed as an indirect measure since it takes into account several variables that have to be measured, as reflected in the equations for the determination of this property, such as the tension applied on the specimens, their length and diameter, etc.

Once the different contributions to uncertainty were analysed, finally three groups were selected:

- Uncertainties associated to testing equipment
- Uncertainties associated with the repeatability of measurements
- Uncertainties associated with the reproducibility of measurements

Of these, the one that brings the greatest contribution is reproducibility, what is consistent with the characteristics of the samples and the test method itself, where various steps are involved to get the final result.

To facilitate the calculations, it has also developed an Excel spreadsheet² in which technicians just have to include the specific data of each laboratory for each of the test methods.

6. ROUND ROBIN TEST

Coming up next, the most important data of the RRT conducted by 15 laboratories from Aleas network regarding the water sensitivity of AC and BBTM type B asphalt mixtures:

6.1. Test program

The tests were conducted in this RRT according to EN 12697-12: 2009. Should be noted that a testing protocol was supplied to each participant in order to minimize any differences in the performance, as well as a spreadsheet to calculate results. The test was carried out on two sets of 4 samples and two types of bituminous mixtures: AC type and another type BBTM B.

6.2. Sampling

To conduct this RRT the collaboration of the company PADECASA, supplying mixture BBTM 11B 45 / 80-65 (corneal aggregate) and the company CAMPEZO supplying an AC16 surf 50/70 S mixture (ophitic-limestone / slag aggregates).

6.3. Determination of test precision. Repeatability and reproducibility.

To determine the precision values, a statistical study was carried out following the UNE 82009-2 and UNE 82009-6 (equivalent to ISO 5725-2 and ISO 5725-6, respectively), which describe a basic method for the determination of repeatability and reproducibility of a standard measurement method. By Through this study, based on data provided by the participants, it is determined whether there are aberrant data and, where appropriate, eliminate them.

In the statistical study the following methods were applied:

a.- Graphical consistency technique. Mandel h and k statistics are used for each laboratory. This is the initial process and with it the inter-laboratory (reproducibility) (value of h), and intra-laboratory (repeatability) (k value) consistency are determined by comparing the values obtained against reference values for the each of participating laboratories and to a certain number of repetitions.

From the results, the differences between laboratories are analysed, as a first approximation of the behavior of each laboratory. These data can be used to determine its relative position and to internally assess the causes that might have produced these differences.

A laboratory qualified at this stage as outlier does not implies its removal in the later study.

b.- Aberrant data detection numerical tests:

² This spreadsheet can be downloaded from the following link <http://www.asefma.es/aleas/>

Two statistical methods have been applied to the study of aberrant values and determining those that should be eliminated from the study.

- Cochran test. Analyzes the standard deviation under repeatability conditions, and to detect laboratories with excessive variability regarding the set of laboratories. This is the initial process wherein a first approximation of the existence of anomalous data.
- Grubbs test (single and double). It detects laboratories presenting results that differ excessively from the average.

In both methods the values obtained are compared with statistical tables on based on 99 and 95% probability of detecting aberrant data.

The criteria for managing test results are as follows:

- If the test statistic is less than or equal to 5% of its critical (greater than or equal in the case of double Grubbs) value, the verified value is accepted as correct.
- If the test statistic is greater than or equal to 5% of its critical (or smaller or equal in the case of double Grubbs) value, and less than or equal to 1% of its critical value (greater than or equal in the case of double) Grubbs, the checked value is considered anomalous, but not eliminated for estimating the overall mean and variances.
- If the test statistic is greater than or equal to 1% of their critical (or smaller or equal in the case of double Grubbs value), the checked value is considered statistically inconsistent and is removed to estimate the overall mean and variances.

To rule out a value, it will require that both, graphical consistency techniques and aberrant detection tests, agrees.

The average value the variance of intra-laboratory reproducibility, the variance of inter-laboratory reproducibility and total reproducibility associated method are calculated using values considered as acceptable. Calculation procedures used are those described in the standard UNE 82009-2. The repeatability and reproducibility values were calculated from the corresponding standard deviations by the following equations:

$$r = 2 \cdot \sqrt{2} \cdot s_r$$

$$R = 2 \cdot \sqrt{2} \cdot s_R$$

Recently, it is considered more appropriate using robust statistics to deal with RRT, mainly when, as in our case, the number of participants is not very high. Therefore the methodology described in the ISO 13528 standard is going to be applied on data of this exercise and the results of both methodologies, ISO 13528 and UNE 82009, will be compared.

Statistics described in ISO 13528 are based on the properties of the median. The assigned value is determined by the robust mean of the results, calculated according to the algorithm A. The standard deviation of the assigned value, used to evaluate the results of the participants, coincides with the robust standard deviation also calculated according to the algorithm A. The algorithm A system is based on an iterative process until data convergence is achieved. The detected anomalous values are not removed but replaced by calculated extreme values. Should be noted that this methodology cannot assess intra-laboratory precision, therefore, the results detected as anomalous will be due to inter-laboratory precision (values differ more than expected respect to the assigned value). So they cannot be compared with those obtained with Cochran test or k-Mandel, but with Grubbs test results or h-Mandel.

A summary of the values obtained in the RRT are shown in Tables 1 and 2:

Table 1: Values of the standard deviation of repeatability and reproducibility for each feature of the AC mixture compared with those given in the relevant test standards.

TEST	P	m	REFERENCE DATA ACCORDING TO STANDARDS		INTERLABORATORY DATA (UNE 82009)		INTERLABORATORY DATA (ISO 13528)	
			S _r (repeatability)	S _R (reproducibility)	S _r (repeatability)	S _R (reproducibility)	m	S _R
ITS (dry) MPa	15	2,249	0,346	-	0,102	0,235	2,262	0,204
ITS (wet) MPa	15	2,126	0,315	-	0,080	0,264	2,189	0,152
ITSR (%)	15	94,5	5,30	8,13	4,9	6,6	94,4	3,6

Table 2: Values of the standard deviation of repeatability and reproducibility for each feature of the BBTM mixture compared with those given in the relevant test standards.

TEST	P	m	REFERENCE DATA ACCORDING TO STANDARDS		INTERLABORATORY DATA (UNE 82009)		INTERLABORATORY DATA (ISO 13528)	
			S _r (repeatability)	S _R (reproducibility)	S _r (repeatability)	S _R (reproducibility)	m	S _R
ITS (dry) MPa	15	1,76	0,346	-	0,085	0,246	1,768	0,225
ITS(wet) MPa	15	1,63	0,315	-	0,075	0,219	1,626	0,185
ITSR (%)	15	92,9	5,30	8,13	5,57	6,66	94,4	3,6

If S_R data are expressed as a coefficient of variation (% CV), the following values are obtained:

Table 3. Coefficients of variation of ITS and ITSR tests

	%CV _R (UNE 82009)	%CV _R (ISO 13528)
ITS wet	12	9
ITS dry	10	7
ITSR	7	4

That means that % CV values around 10% are obtained for ITS, and lower for the ITSR. This leads us to believe that there may be potential systematic errors in ITS test, which are compensated when calculating the value of ITSR.

6.4 Estimation of uncertainty.

Taking advantage of this RRT, we proceeded to the estimation of measurement uncertainty of water sensitivity test (resistance to indirect tensile test under wet and dry conditions and retained strength). This uncertainty has been estimated based on equation showed in the standard, to reach a value of global uncertainty. To do this we proceeded to evaluate the influence that each of the terms that take part of this equation, representing a physical magnitude,

Contributing physical magnitudes within the test are: weighting, dimensions, and temperature and load measurements. These figures, in the study of quantifying uncertainty in turn are complemented with coefficients indicating the participation of each in the final calculation.

Besides taking into account the physical parameters of the test, there are other measurable variables that influence the estimate of overall uncertainty relating to the repeatability and reproducibility of different measurements.

Bearing in mind the above considerations, some spreadsheets that have been used to estimate the overall uncertainty values associated with the results of this RRT.

To do this they, confidential data provided by the laboratories have been used, Regarding uncertainties associated to the calibration of equipment used in various measurements to be carried out for the development of water sensitivity test according to EN 12697-12 standard., the devices that have been taken into account in the calculations of the uncertainties are as follows: balance, gauge, thermometer and compression testing machine.

In the following table, the average values, maximum and minimum values of the expanded uncertainties brought are summarized, by the different laboratories for each of the equipment used in this RRT.

Table 4: Overall average of expanded uncertainties for each piece of equipment in the (between brackets maximum and minimum values are shown.

Equipment	Expanded uncertainty
Compression machine (KN)	0,074 (0,22-0,007)
Gauge (mm)	0,05 (0,2-0,008)
Balance (g)	0,21 (0,71-0,003)
Thermometer (°C)	0,45 (1,6-0,12)

To estimate the uncertainty of this exercise the average values of different measured variables were taken, as well as the average values of the uncertainties of the equipment. The results obtained for indirect tensile strength and water sensitivity of each mix type are presented in Tables 5 and 6.

Table 5: Uncertainty values of BBTM B mixture

Test	Assigned value	Expanded uncertainty	% relative extended uncertainty
IT dry (MPa)	1,76	0,42	23,9
IT wet (MPa)	1,63	0,51	31,3
ITSR (%)	92,9	40	43,1

Table 6: Uncertainty values for AC mixture

Test	Assigned value	Expanded uncertainty	% relative extended uncertainty
IT dry (MPa)	2,249	0,44	19,6
IT wet (MPa)	2,126	0,51	24,0
ITSR (%)	94,5	29	30,7

The relative expanded uncertainty for each test is calculated by dividing the expanded uncertainty by the assigned value, expressed in%.

As for the results of uncertainties of water sensitivity test, it can be concluded:

- In relation to the uncertainty values obtained for the ITSR for both mixtures, one must not forget that the estimated values correspond to those obtained in the RRT. These data can be used as a reference but it is important to note that each laboratory should confirm whether their estimated uncertainty values have the same order of magnitude.
- In any case, applying uncertainty methodology developed by Aleas group, values of relative uncertainty ranging 20-30% were obtained for ITS tests and 30-40% for ITSR (see Tables 10 and 11).
- In ITS testing we see that the reproducibility is the most important source of uncertainty (over 80%), so a better value CVR should result in better value uncertainty contribution. For ITSR test, since it is a simple calculation with the values of ITS, CVR better values are obtained than those obtained for ITS, which should lead to lower values of uncertainty. However, the strict application of the mathematical method to estimate uncertainty translates into higher values, so that we may be overestimating ITSR uncertainty. A more realistic estimate, based on the assumption $I \sim 2sR$ (applicable where, as in our case, the main contribution to the uncertainty is due to random factors, such as the reproducibility of the assay), leads to relative values of ITSR uncertainty ranging 15%.

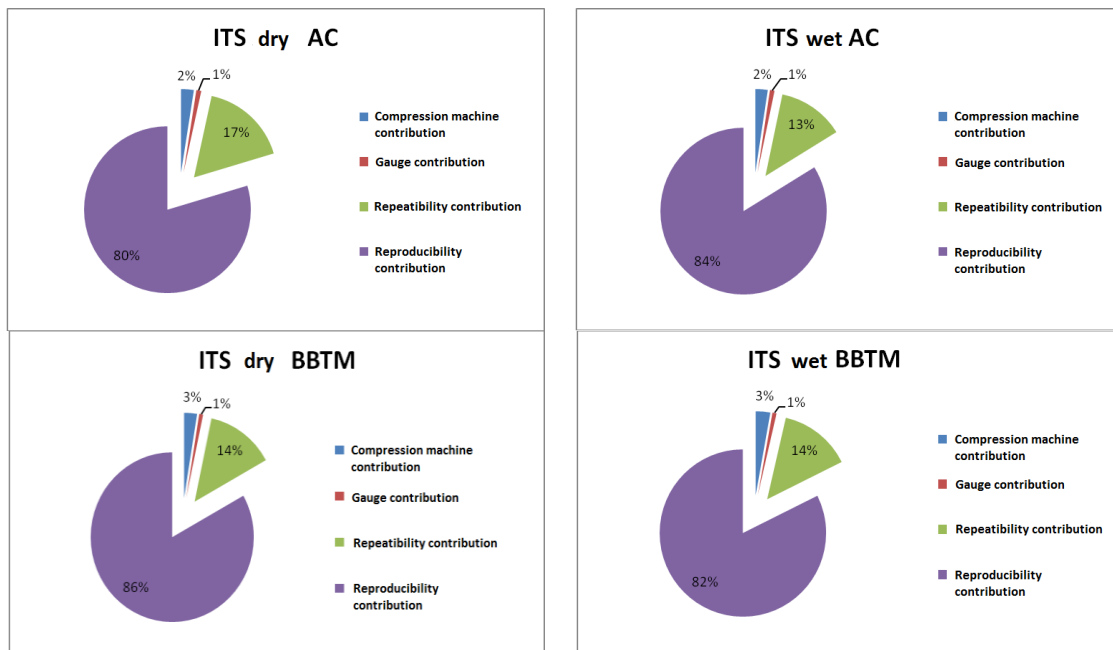


Figure 3: Graphic representation of uncertainty contributions in water sensitivity test (EN 12607-12)

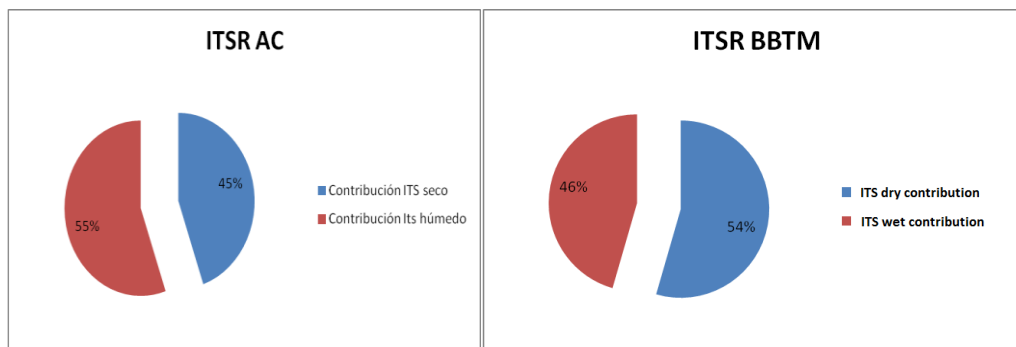


Figure 3: Graphic representation of uncertainty contributions of ITS (dry and wet on) ITSR test (EN 12607-12)

7. ANALYSIS OF RESULTS

Once quantified the uncertainty of the results of water sensitivity test of asphalt mixtures, we will proceed to analyse which implications arise from its application to product characteristics defined in EN 13108-1 and Spanish specifications (PG-3).

Firstly, ITSR categories of AC mixes detailed in EN 13108-1 are classified in 10% ranges, from 60% up to a maximum of 90%. In the case of the Spanish specifications, the required categories are 80% for basic and intermediate layers and 85% for wearing courses.

This study, focused on the uncertainty estimate in EN 12697-12 standard, shows uncertainty values of 29%, what means that an actual value of a measurement of sensitivity water defined as 80%, should be completed adding its uncertainty: that is $80 \pm 29\%$. That means that any value between 51% and 109% should be valid. Two questions arise from these results:

- The whole ranges defined in EN 13108-1 and EN 13108-2 for water sensitivity test are overlapped by the uncertainty of this experimental method.
- Provided that testing a value of 80% has been obtained, should it be considered as acceptable by Road Administration, which requires a nominal value of 90%, since both values are within the extended range of uncertainty?

From a metrological point of view, a result of 80% should comply with the specifications requested since, according to the methodology used; any value between 51% and 109% should be accepted as valid. Therefore, are specifications consistent with the uncertainty provided by this testing?

On the other hand, regarding the ranges specified for each category of EN 13108-1 and EN 13108-2, are consistent these divisions as well?

Results showed in this RRT are similar to those published in the NCHRP Project 9-26^a [3] report, which found values between 15 and 30% of CVR in ITS test, for specimens prepared with Marshall compactor and gyratory compactor. As in this case, uncertainty CVR values are reduced for ITSR (up to 10%).

All these results generate doubts about the suitability of the test method EN 12697-12 (A), since no longer allows to properly differentiate changes in a bituminous mixture generated by a change of aggregates, additives, manufacturing conditions etc. Therefore it is necessary to analyse the various sources of uncertainty for minimizing them. In this regard, most of the standards EN 12697 have severe deficiencies in defining the precision requirements for test equipment, mixing and misuse terms such as resolution, precision and accuracy. Proper preparation and use of these concepts might allow a reduction of the overall uncertainty of the tests. It would also be particularly useful to the standards practitioners including a detailed bibliographic reference about repeatability and reproducibility studies, in order to provide consistent clues about the order of magnitude of test variability.

8. CONCLUSIONS

- Uncertainty obtained in water sensitivity test EN 12697-12 invalidate using categories specified in the of the Standards EN 13108-1 and EN 13108-2, since uncertainty is larger than the ranges specified for each category
- It was not possible to identify the most relevant factors that have generated the above mentioned uncertainty values, so it will be necessary to continue the experimental work to narrow and adequately reduce this uncertainty.
- The uncertainty calculation methodology used in this work allows the automation of calculations, making it easy for laboratories self-control tasks.
- It has been observed that many of the standards EN 12697 does not adequately define the elements that control the precision of the tests, so the possibilities of generating high levels of uncertainty increase substantially.

9. ACKNOWLEDGEMENTS

The authors wish to thank all members of Aleas who participated in the round robin test their generous contribution to the development and the experimental data analysis.

10. BIBLIOGRAPHIC REFERENCES

- [1] International Vocabulary of Metrology. Basic and general concepts and associated terms (VIM). 3rd edition. 2012
- [2] Guía para la estimación de la incertidumbre en métodos de ensayo de control de mezclas bituminosas. Asefma. 2015.
- [3] NCHRP Project 9-26A, September 2010 Precision Estimates of AASHTO T283 Resistance of Compacted Hot Mix Asphalt to Moisture-Induced Damage