B

STATISTICAL ISSUES (COST 323 SPECIFICATION)

There are numerous applications for WIM systems each requiring a particular configuration and level of accuracy of measurement. To be able to evaluate and compare the performance of these systems, it is necessary to define criteria for evaluation or for acceptance. Jacob and O'Brien (1997, 1998) describe the philosophy behind the European WIM specification, prepared by the COST323 management committee. This specification gives an indication of what accuracy might be achievable from sites with particular characteristics, and what accuracy might be acceptable for various needs (COST 323 1997). Calibration and testing procedure for WIM systems are considered. Accuracy classes are defined on the basis of the width of the confidence interval within the measured results lie. The specification defines seven accuracy classes, A(5), B+(7), B(10), C(15), D+(20), D(25) and E. For each class, confidence interval widths are specified for gross weights, weights of axles, axle groups, speed, inter-axle distance and axle/vehicle count. The required level of confidence is a function of the test conditions and the number of test runs.

The WIM system accuracy classification is based on comparisons of measured results against reference values which, it is anticipated, would generally be determined by statically weighing trucks. To comply with a given accuracy class, the calculated probability that individual results (i.e. static load W_s) are within the confidence interval $[W_s(1-\delta), W_s(1+\delta)]$, or that individual relative errors are within the confidence interval $[-\delta, +\delta]$, must exceed a specified minimum, π_0. The confidence interval width, δ, is a function of the accuracy class and the specified values are given in Table B.1 for each identity.

Table B.1 – Tolerances of the Accuracy Classes (Confidence Interval Width for Relative Errors)

Type of measurement	Domain of use	Confidence interval width d(%) for Each Accuracy Class					
		A(5)	**B+(7)**	**B(10)**	**C(15)**	**D(25)**	**E**
Gross weight	Greater than 35 kN	5	7	10	15	25	>25
Group of axles	Greater than 20 kN	7	10	13	18	28	>28
Single axle	Greater than 20 kN	8	11	15	20	30	>30
Axle of a group	Greater than 20 kN	10	15	20	25	35	>35
Speed	Greater than 30 km/h	2	3	4	6	10	>10
Inter-axle distance		2	3	4	6	10	>10
Axle/vehicle count		1	1	1	3	5	>5

Three main types of application are identified for road authorities, decision makers and WIM users (COST323 1997):

- Statistics: Economical and technical studies of freight transport, general traffic evaluation on roads and bridges, collecting statistical data, etc.. d in the range of 15 to 30% (class C(15), or D(25)).
- Infrastructure and preselection: Detailed analysis of traffic, design and maintenance of roads and bridges, accurate classification of vehicles, preselection for enforcement, etc.. d in the range of 10 to 20% (class B(10), or C(15)).
- Enforcement or legal weight limits: Enforcement, commercial and industrial applications. d: 5 to 10% (class A(5), or B+(7)).

B.1 CALIBRATION

Pre-weighed trucks are preferred for calibration of WIM systems. The specification defines four levels of repeatability/reproducibility test conditions as follows:

- Full Repeatability Conditions (r1): One vehicle passes several times at the same speed, load and lateral position.
- Extended Repeatability Conditions (r2): One vehicle passes several times at different speeds, different loads and with small variations in lateral position (in accordance with typical traffic).

- Limited Reproducibility Conditions (R1): A small set of vehicles (typically 2 to 10), representative in weight and silhouette of typical traffic, is used. Each vehicle passes several times, at different combinations of speed and load and with small variations in lateral position.
- Full Reproducibility Conditions (R2): A large sample of vehicles (some tens to a few hundred), taken from the traffic flow and representative of it, is used for calibration.

Calibration under full repeatability conditions is not recommended unless the customer approves. The vehicle should be typical of the traffic being weighed and should be driven in fully loaded, half-loaded and eventually unloaded conditions, and at speed levels representative of the site. Calibration under limited reproducibility (or extended reproducibility) conditions is recommended. The set of calibration vehicles and loading cases should be representative of the traffic pattern. In an initial verification, in which the same sample is used for calibration and for accuracy analysis, to be accepted in an accuracy class defined by $\boldsymbol{d}$, the specification requires that the confidence interval becomes $[-k.\boldsymbol{d},+k.\boldsymbol{d}]$ for the relative errors at a level of confidence greater than $\boldsymbol{p_0}$. If the calibration is made using repeated runs of pre-weighed trucks, k is taken equal to be 0.8. This value of k gives the best agreement between the accuracy results for a system from the initial verification check and the in-service check.

B.2 ACCURACY CLASSIFICATION FOR IN-SERVICE WIM SYSTEMS

Repeated runs of pre-weighed test vehicles, and/or the use of single runs of pre- or post-weighed vehicles from the traffic flow can be used to assess the accuracy of an operational WIM system. When the in-service accuracy of a system must be verified, the test may be carried out during various time periods and environmental repeatability or reproducibility conditions are defined:

- Environmental Repeatability (I): The test time period is limited to a few hours such that the temperature, climatic and environmental conditions do not vary significantly during the measurements.
- Limited Environmental Reproducibility (II): The test time period extends over at least 24 hours period, or preferably over a few days within the same week or month, such

that the temperature, climatic and environmental conditions vary during the measurements, but no seasonal effect has to be considered.

- Full Environmental Reproducibility (III): The test time period extends over a whole year or more, or at least over several days spread all over a year, such that the temperature, climatic and environmental conditions vary during the measurements and all the site seasonal conditions are encountered.

The specification does not allow recalibration of the system with the sample of data collected for the in-service test, and therefore the eventual bias must not be removed. It is recommended to perform the check in conditions (R1) or (R2). It may be done in conditions (r2) but with at least 3 loading conditions uniformly distributed within the range of axle/gross weights to be weighed, and 10 runs per loading case. It is not recommended to perform the check in conditions (r1), unless special agreement of the user. After such an in-service verification, and if a large bias is found which leads to a lower accuracy class than expected, a recalibration may be carried out.

The minimum probability, π_0, is a function of the sample size *n* and the test conditions (repeatability or reproducibility (r1) to (R2), environmental variability (I) to (III)). The minimum values π_0 of the required level of confidence for the centered confidence intervals specified in Table B.1 are given in Tables B.2 to B.4. For sample size not mentioned in these Tables, the figures may be interpolated.

Table B.2 – Minimum levels of confidence π_0, of the centred confidence intervals (in %) case of a test under 'Environmental Repeatability' (I)

Sample size (*n*) **Test conditions**	**10**	**20**	**30**	**60**	**120**	**¥**
Full repeatability	*95*	*97.2*	*97.9*	*98.4*	*98.7*	*99.2*
Extended repeatability	*90*	*94.1*	*95.3*	*96.4*	*97.1*	*98.2*
Limited reproducibility	*85*	*90.8*	*92.5*	*94.2*	*95.2*	*97.0*
Full reproducibility	*80*	*87.4*	*89.6*	*91.8*	*93.1*	*95.4*

Table B.3 – Minimum levels of confidence π_0, of the centred confidence intervals (in %) case of a test under 'Limited Environmental Reproducibility' (II)

Sample size (*n*) **Test conditions**	**10**	**20**	**30**	**60**	**120**	**¥**
Full repeatability	*93.3*	*96.2*	*97.0*	*97.8*	*98.2*	*98.9*
Extended repeatability	*87.5*	*92.5*	*93.9*	*95.3*	*96.1*	*97.5*
Limited reproducibility	*81.9*	*88.7*	*90.7*	*92.7*	*93.9*	*96.0*
Full reproducibility	*76.6*	*84.9*	*87.4*	*90.0*	*91.5*	*94.3*

Table B.4 – Minimum levels of confidence π_0, of the centred confidence intervals (in %) case of a test under 'Full Environmental Reproducibility' (III)

Sample size (*n*) **Test conditions**	**10**	**20**	**30**	**60**	**120**	**¥**
Full repeatability	*91.4*	*95.0*	*96.0*	*97.0*	*97.6*	*98.5*
Extended repeatability	*84.7*	*90.7*	*92.4*	*94.1*	*95.1*	*96.8*
Limited reproducibility	*78.6*	*86.4*	*88.7*	*91.1*	*92.5*	*95.0*
Full reproducibility	*73.0*	*82.3*	*85.1*	*88.1*	*89.8*	*93.1*

B.3 PROCEDURE FOR THE ASSESSMENT OF ACCURACY OF A WIM SYSTEM

The relative errors with respect to the static weights and loads (or any other accepted reference values) are calculated, for each measurement of gross weight, single axle load, etc., as follows:

$$x_i = \frac{Wd_i - Ws_i}{Ws_i} \tag{B.1}$$

where Wd_i and Ws_i are the in-motion measured value and the reference (static) value respectively on the same unit. Then, the sample statistics to be calculated and used to classify the system are the mean *m*, the standard deviation *s* and the number of values, *n* of all the measured values (x_1, x_2, ..,x_n) for each identity.

The confidence level p may be either estimated by a theoretical method using the sample statistics of the test, or, in some cases, by a sample proportion. An individual value is considered of a relative error, taken randomly from a Normally distributed sample of size *n*, with a sample mean m and standard deviation *s*. A lower bound p, on the probability for that individual value to be in the centered confidence interval [$-\boldsymbol{d}$,$+\boldsymbol{d}$], is given at the confidence interval *(1-$\boldsymbol{a}$)* by:

$$\boldsymbol{p} = \Phi(u_1) - \Phi(u_2) \tag{B.2}$$

where

$$u_1 = \frac{(\boldsymbol{d} - m)}{s} - \frac{t_{v,1-\boldsymbol{a}/2}}{\sqrt{n}} \tag{B.3}$$

$$u_2 = \frac{(-\boldsymbol{d} - m)}{s} + \frac{t_{v,1-\boldsymbol{a}/2}}{\sqrt{n}} \tag{B.4}$$

and Φ is the cumulative distribution function of a Student variable, and $t_{v,1-\boldsymbol{a}/2}$ is a Student variable with $v=n-1$ degrees of freedom. The parameter α is taken to equal 0.05. Then the estimated level of confidence $\boldsymbol{p}$, for each sample (and criterion) is calculated using Equation B.2. For the case of an initial verification (same data used for calibration and checking), $\boldsymbol{d}$ is replaced by $k.\boldsymbol{d}$ in Equation B.2.

The accuracy level of the WIM system is assessed using one of two methods:

- For each sub-population (sample) corresponding to a criterion from Table B.1, and for the proposed (required) accuracy class defined by $\boldsymbol{d}$, the acceptance test is:
  a) If $\boldsymbol{p} \geq \boldsymbol{p}_0$, the system is accepted in that class;
  b) If $\boldsymbol{p} < \boldsymbol{p}_0$, the system can not be accepted in the proposed accuracy class, and the acceptance test is repeated with a lower accuracy class (a greater $\boldsymbol{d}$). $\boldsymbol{p}$ should be recalculated using Equation B.2.
- An alternative way is to calculate, using Equation B.2, the (lowest) value of $\boldsymbol{d}$ which provides: $\boldsymbol{p} = \boldsymbol{p}_0$, and then to check that $\boldsymbol{d}$ is less than value specified in Table B.1 for the proposed accuracy class and criterion. This approach allows a system to be classified in any accuracy class defined by an arbitrary $\boldsymbol{d}$-value (the best one). It is mainly recommended for detailed research and testing studies.

Table B.5 shows the result of an accuracy classification for given sample statistics. It can be seen that the WIM system fulfils the requirements of class B+(7), and, for the single axle, group of axles and gross weight criterions, class A(5). The d_{min} value refers to the lowest d value that gives a theoretical p value equal to the required p_0 value. Accordingly, gross weight is the most accurate criterion in this classification.

Table B.5 – Accuracy classification

(**n**: Total number of vehicles; **m**: mean; **s**: Standard deviation; $\mathbf{p_0}$: level of confidence; **d**: tolerance of the retained accuracy class; $\mathbf{d_{min}}$: minimum width of the confidence interval for π_0; **p**: Level of confidence of the interval [-δ,δ])

	Relative error statistics				Accuracy calculation				
Criterion	**n**	**m (%)**	**s (%)**	$\mathbf{p_o}$ **(%)**	**Class**	**d (%)**	$\mathbf{d_{min}}$ **(%)**	**p (%)**	**Class Retained**
Single axle	27	0.50	2.07	92.1	**A(5)**	8	4.7	99.8	**B+(7)**
Axle of group	18	-0.09	5.39	90.3	**B+(7)**	14	12.1	94.9	
Group of axles	9	-0.45	0.52	83.4	**A(5)**	7.14	1.4	100.	
Gross Weight	18	0.00	0.30	90.3	**A(5)**	5	0.7	100.	

B.4 CHOICE OF WIM SITE

The site characteristics can have a significant influence on the performance of WIM systems. A rough road surface induces high dynamic forces leading to inaccurate results. Deflection, rutting and pavement cracking can cause sensor failure. So, WIM sites are divided in three classes according to rutting, deflection and evenness (Table B.6). Table B.7 indicates the accuracy class that is likely to be achievable at a site of a given class.

The accuracy of a bridge WIM system also depends highly on the selection of the weighing site, particularly on the type of the superstructure and the evenness of the approach.

Table B.6 – Classification and Criteria of WIM sites

			WIM Site Class		
			I Excellent	**II Good**	**III Acceptable**
Rutting (3 m – beam)		Rut depth max. (mm)	≤ 4	≤ 7	≤ 10
Deflection (quasi-static) (130 kN – axle)	Semi-rigid pavements	mean deflection (10^{-2} mm)	≤ 15	≤ 20	≤ 30
		left/right difference (10^{-2} mm)	± 3	± 5	± 10
	All bitumen pavements	mean deflection (10^{-2} mm)	≤ 20	≤ 35	≤ 50
		left/right difference (10^{-2} mm)	± 4	± 8	± 12
	Flexible pavements	mean deflection (10^{-2} mm)	≤ 30	≤ 50	≤ 75
		left/right difference (10^{-2} mm)	± 7	± 10	± 15
Deflection (dynamic) (50 kN – load)	Semi-rigid pavements	deflection (10^{-2} mm)	≤ 10	≤ 15	≤ 20
		left/right difference (10^{-2} mm)	± 2	± 4	± 7
	All bitumen pavements	mean deflection (10^{-2} mm)	≤ 15	≤ 25	≤ 35
		left/right difference (10^{-2} mm)	± 3	± 6	± 9
	Flexible pavements	mean deflection (10^{-2} mm)	≤ 20	≤ 35	≤ 55
		left/right difference (10^{-2} mm)	± 5	± 7	± 10
Evenness	IRI index	Index (m/km)	0 – 1.3	1.3 – 2.6	2.6 – 4
	APL[(1)]	Rating (SW,MW,LW)[(1)]	9 - 10	7 - 8	5 - 6

[(1)] *The APL is a car-towed device used to measure longitudinal profile in the short (SW), medium (MW) and long wavelengths (LW) respectively.*

Table B.7 – WIM Accuracy Class Likely to be Achievable in given WIM Site Class

Accuracy Class	WIM Site Class		
	I (Excellent)	II (Good)	III (Acceptable)
Class A(5)	Sufficient	Insufficient	Insufficient
Class B+(7)	Sufficient	May be sufficient	Insufficient
Class B(10)	Sufficient	Sufficient	Insufficient
Class C(15)	More than sufficient	Sufficient	Sufficient
Class D+(20)	More than sufficient	More than sufficient	Sufficient
Class D(25)	More than sufficient	More than sufficient	Sufficient