NORDTEST Report

TR 533
Approved 2003-04

Inter-laboratory comparison of fatigue test with evaluation of the participating laboratories calculations of measurement uncertainty

Thomas Svensson
Magnus Holmgren
Klas Johansson
Erland Johnson

Published by Nordtest
Tekniikantie 12
FIN–02150 Espoo
Finland

Phone: + 358 9 455 4600
E-mail: nordtest@nordtest.org

Fax: + 358 9 455 4272
Internet: www.nordtest.org

NT TECHN REPORT 533
Approved 2003-04

Authors: Thomas Svensson [1] Magnus Holmgren [2] Klas Johansson [2] Erland Johnson [2]	**NORDTEST project number:** 1591-02
	Institution: [1)] Fraunhofer-Chalmers Research Centre for Industrial Mathematics [2)] SP Swedish National testing and research Institute

Title (English):	-
Title (Original):	Inter-laboratory comparison of fatigue test with evaluation of the participating laboratories calculations of measurement uncertainty

Abstract:

In this paper a fatigue testing inter-laboratory comparison is presented. Six Nordic laboratories performed fatigue tests on steel specimens. The specimens were tested with different stress levels (460 MPa, 430 MPa and 400 MPa) and with the load ratio $R=S_{min}/S_{max}=0.1$. On each level four specimens were tested. The participating laboratories reported the results together with measurement uncertainty.

The results show a significant difference between laboratories. However, when the differences in modelling were taken into consideration no significant differences between the laboratories remained. There are large differences in the way the measurement uncertainties were calculated and reported. No laboratory did take the most influencing uncertainty parameter, bending stress (due to misalignment of the testing machine, “incorrect” specimens and incorrectly mounted specimens), into consideration when calculating the measurement uncertainty.

Technical Group: Expert Group Quality and Metrology		
ISSN: 0283-7234	**Language:** English	**Pages:** 20
Class (UDC): 620.1	**Key words:** Inter laboratory comparison, fatigue test, labaratories, calculation, uncertainty	
Distributed by: **NORDTEST** **Tekniikantie 12** **FIN-02150 ESPOO** **Finland**	**Publication code:**	
	Report Internet address: http://www.nordtest.org/register/techn/tlibrary/tec533.pdf	

Inter-laboratory comparison of fatigue test with evaluation of the participating laboratories calculations of measurement uncertainty (Nordtest-project 1591-02)

Thomas Svensson (Fraunhofer-Chalmers Research Centre for Industrial Mathematics, Magnus Holmgren, Klas Johansson and Erland Johnson (SP Swedish National testing and research Institute)

ABSTRACT

In this paper a fatigue testing inter-laboratory comparison is presented. Six Nordic laboratories performed fatigue tests on steel specimens. The specimens were tested with different stress levels (460 MPa, 430 MPa and 400 MPa) and with the load ratio $R = S_{min}/S_{max} = 0.1$. On each level four specimens were tested. The participating laboratories reported the results together with measurement uncertainty.

The results show a significant difference between laboratories. However, when the differences in modelling were taken into consideration no significant differences between the laboratories remained. There are large differences in the way the measurement uncertainties were calculated and reported. No laboratory did take the most influencing uncertainty parameter, bending stress (due to misalignment of the testing machine, "incorrect" specimens and incorrectly mounted specimens), into consideration when calculating the measurement uncertainty.

INTRODUCTION

Fatigue of metals is a large engineering problem and it is the main cause of many mechanical failures. To avoid such problems, accurate fatigue design of products is therefore very important. To be able to perform a "correct" fatigue design the designer needs reliable material data. However, reliable fatigue data is hard to get, especially since fatigue testing is a problematic area. Therefore quality assurance activities for fatigue testing are important. One such quality assurance activity is inter-laboratory comparisons. Inter-laboratory comparisons help the laboratories to improve their performance as well as to give the customers information about how reliable fatigue data are.

One specific problem, which has arisen through the introduction of the new standard ISO 17025 [1] where it is stated that a test result should be reported with measurement uncertainty, is the calculation of measurement uncertainty of fatigue test results. The way one should calculate and report measurement uncertainty for fatigue tests is not well known.

Therefore, an inter-laboratory comparison including measurement uncertainty calculations has been performed and is reported in this report.

PARTICIPANTS

Six laboratories participated in the inter-laboratory comparison, namely:

VOLVO Technical Development: an industrial laboratory in Sweden

SEMCON: a laboratory in a consultancy company active in Sweden

FORCE Institute: a research institute in Denmark

Laappenranta Technical University: a university laboratory in Finland

Linköping Technical University: a university laboratory in Sweden

SP Swedish National Testing and Research Institute: a research institute in Sweden

Each participant was randomly given a number between 1 and 6 and this notification will be used in the rest of this paper.

Two of the laboratories had an accreditation for fatigue testing

EXPERIMENTAL

Test programme and instructions to the participants

The participants received information about the test specimens (without material data) together with the following instructions:

Performance and results of tests

General

The tests shall be performed:

- As constant load amplitude test, with R = 0.1 ($R= F_{min}/ F_{max}$).
- At three different stress levels, 460 MPa, 430 MPa and 400 MPa.
- With four specimens at each stress level.
- With test frequency between 10 and 30 Hz.
- With run out limit at 5 million cycles.
- In normal laboratory climate (20°C±3°C and 50%±15% RH)

The test results shall be used to calculate estimates of the two fatigue strength parameters A and B (see below). Each estimated parameter contains two types of errors, namely 1) a random error because of the scatter in the material properties and 2) a measurement error because of uncertainties in the measurement procedures. The second type of error is here denoted as the uncertainty in measurement and shall be evaluated by each laboratory according to its own measurement devices. The final result shall include both the estimated parameters A and B and the uncertainties in them due to measurement errors. The report shall also include the considerations and calculations behind the results.

Test results

The following properties for each specimen will be reported (defined in accordance with ASTM E466 [2]): specimen name, thickness (mm), width (mm), maximum and minimum stress (MPa), maximum and minimum load (kN), number of cycles (N) and any remarks.

The properties for the test batch will be reported as:

- Material parameters A and B according to linear regression of the $\log(\sigma)$ –$\log(N)$ curve, $\log(\sigma)=A+B\log(N)$.

- Measurement uncertainty for the results.
- Considerations and calculations behind the measurement uncertainty

Together with the results a description of how the measurement uncertainty was calculated should be provided.

The tests should be performed in accordance with ISO 5725-2 [3], including

- Any preliminary checking of equipment shall be as specified in the standard method for your laboratory.
- The specimens shall be tested within a short interval of time and by the same operator, and without any intermediate recalibration of the apparatus.

If it is impossible to conform to these requirements, each disagreement must be reported.

A staff member of each of the participating laboratories is responsible for organizing the actual performance of the measurements, and for reporting the test results. This supervisor shall

a) Ensure that the operators selected are those who would normally carry out such measurements in routine operations
b) Hand out the samples to the operator in keeping with the specific instructions of the test performance.
c) Collect the test results, including any anomalies and difficulties experienced, and comments made by the operator

The supervisor of each laboratory shall also write a full report that shall contain the information asked for in ISO 5725-2, e.g.: the test results, the originally observed values of readings from which the test results were derived, information about irregularities or disturbances that may have occurred, including any change of operator that may have occurred, together with a statement as to which measurements were performed by which operator, and the reasons for missing results, etc.

Test specimens

The test specimens were produced at the mechanical workshop at SP. For each of the participating laboratory, 12 specimens were manufactured. The test specimens were cut from machined SS 1650 steel plates. Each specimen was marked with an individual number. This number was used when reporting the results. The test specimens were randomly distributed to the participants.

Table 1. Nominal specimen dimensions

Thickness [mm]	Width, test section [mm]	Parallel length [mm]	Width, gripped end [mm]	Radius between test section and gripped end [mm]	Total length [mm]
6	26	65	40	208	300

Table 2. Nominal material data for SS 1650

Yield stress	Tensile strength	Hardness Vickers
375-390 MPa	670-690 MPa	195.3±1.6

The yield stress and the tensile strength are tabulated data while the hardness is based on one measurement. After the tests, the specimens were returned to SP.

Equipment and test

The equipment used by the participating laboratory is listed in table 3.

Table 3. Test equipment, test frequency and preparations

Laboratory	**Equipment (Testing machine, frame, load cell, actuator, test programme)**	**Testing frequency**	**Preparations (alignment, calibration)**
1	Servo hydraulic testing machine, Instron 1341, 100 kN, 8500 plus control	11-13 Hz	Calibrated October 17 2001
2	Two testing machines were used: 100 kN MTS Class 0.5 500 kN MTS Class 0.5	10-15 Hz (two specimens were tested with 15 Hz and two with 10 the rest with 12 Hz)	
3	150 kN double direct servo hydraulic actuator	One test with 2 Hz, three with 5 Hz and the rest with 15 Hz	Alignment checked according to ASTM E466 with instrumented test specimen, bending stress less than 3%
4	Servo hydraulic Instron 8801 and 8800 control Max version 6.7	10 Hz	Align Pro INSTRON was used bending stress lower than 5% Calibrated before the test
5	Force transmitter: KR 000-0001 MTS Flex Test: Flex-0003 MTS load frame: LR00-0001	10 Hz	
6	No report		

MEASUREMENT UNCERTAINTY CALCULATIONS

One of the objectives with the present investigation was to compare the observed differences between laboratory test results with their estimated measurement uncertainty. The uncertainty analyses, as such, were also intended to be studied and compared to the standard procedure recommended in the ISO guide: Guide to the expression of uncertainty in measurements (GUM) [4].

The laboratories have identified different sources of uncertainty and have treated them in different ways. These sources are the load measurement, the load control, the superimposed bending stresses because of misalignment, and the dimension measurements. Implicitly also laboratory temperature and humidity, specimen temperature and corrosion effects are considered. In addition, the results show a modelling effect that we will discuss in the sequel. In table 4 we summarize the different laboratory treatments of these sources.

Table 4. Sources of uncertainty and laboratory treatment

Source	lab 1			lab 2			lab 3			lab 4			lab 5			lab 6		
	C	N	A	C	N	A	C	N	A	C	N	A	C	N	A	C	N	A
Load cell	X		X	X	X		X	-	-	X		X	X		X			
Load control	X		X	-	-	-	X	-	-	X		X	X		X			
Bending stress	-	-	-	-	-	-	X	-	-	X	-	-	-	-	-			
Dimensions	X		X	X	X		-	-	-	X	-	X	-	-	-			
Lab. temp.	X	X		X	X		X	X		X	X		X	X				
Lab. humidity	X	X		X	X		X	X		X	X		X	X				
specimen temp.	X	-	-	X	-	-	-	-	-	-	-	-	X	-	-			
corrosion	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-			
modelling	X	X		X	-	-	-	-	-	-	-	-	-	-	-			
frequency	X	X		-	-	-	-	-	-	-	-	-	-	-	-			
Failure criterion	-	-	-	-	-	-	-	-	-	X	X		-	-	-			

C : The laboratory report considers explicitly or implicitly the source.
N: The laboratory report neglects the source
A: The laboratory report takes the source into account in the measurement uncertainty calculation.

Specific comments on the different laboratories

All laboratories have given their laboratory temperature and humidity, but not considered these values as sources of uncertainty. Since limits for these influences were specified in the instructions they should be regarded as a part of the test method and not included in the uncertainty considerations.

Laboratory 1
The uncertainty in the applied stress has been made taking load cell, load control and dimension uncertainties into account. The mathematical evaluation has been made in accordance with GUM. Specimen temperature has been measured, but is implicitly neglected. The modelling problem is mentioned, but not considered as an uncertainty source.

Laboratory 2
The report contains no uncertainty evaluation. The uncertainties in load cell and micrometer are considered, but neglected with reference to the large material scatter. Specimen temperature has been measured. Modelling problems are mentioned by a comment regarding the choice of load levels.

Laboratory 3
The report contains no uncertainty evaluation. However, the accuracy of the machine is given and the load was controlled during test to be within specified limits. The bending stresses have been measured on one specimen, but its influence on the fatigue result is not taken into consideration.
In a final version of the report, lab 3 included an uncertainty calculation. Unfortunately this version was submitted too late to be analysed for this summary report.

Laboratory 4
The uncertainties in load cell and dimension measurements are considered in an evaluation of stress uncertainty. The method for the evaluation is not in accordance with the GUM method, but is performed by adding absolute errors. The bending stress influence and the control system deviations are considered, but not included in the uncertainty evaluation. Failure criterion is mentioned and regarded negligible and corrosion is mentioned as a possible source of uncertainty.

Laboratory 5
Uncertainties in the load cell and the load control have been considered and the evaluation of the load uncertainty has been performed according to the CIPM-method.

Laboratory 6
No report has been provided, but only experimental results and a Wöhler curve estimate.

No laboratory has reported the uncertainty in the estimated material properties, the Wöhler parameters, but at most the uncertainty in the applied stress. The overall picture of the uncertainty considerations is that only uncertainty sources that are possible to estimate from calibration reports are taken into account in the final evaluation.

One important source that several laboratories have mentioned is the bending stresses induced by misalignment in the testing machine, incorrectly mounted test specimens and "incorrect" specimens. The amount of bending stresses is also estimated in some cases, but its influence on the uncertainty in the final Wöhler curve is not investigated.

The results from this experimental investigation show that there are different ways of determining the Wöhler curve from the experimental result. One problem is the surviving specimens, the run out results. Four laboratories use only the failed specimens results for the curve fit, one laboratory neglects all results on the lowest level, and one laboratory includes the run outs in the estimation. Another problem is the mathematical procedure for estimating the curve. The common practice and the recommendation in the ASTM standard is that the curve should be estimated by minimizing the squared errors in log life, i.e. the statistical model is

$$\lg N = \lg a - b \lg S + \varepsilon \quad , \tag{1}$$

where ε is a random error, assumed to have constant variance and where lg stands for the logarithm with base 10.

One of the laboratories made the estimation in the opposite direction, i.e. the squared errors in logarithmic stress were minimized, which lead to a model discrepancy discussed below.

RESULTS

The primarily laboratory results that should be compared are the estimated Wöhler curves. In order to compare them they need to be written in the same format. The Wöhler curve can be written in different ways. In our analysis we rewrite it with the stress as independent variable. In order to present all results in the same way we need to transform some of them. Two laboratories have given the parameters in the formula

$$\lg S = A - B \lg N$$

The resulting parameters were transformed to formula (1) parameters by

$$\lg a = \frac{A}{B} \quad , \quad b = \frac{1}{B}$$

One problem when comparing two linear curves is that the parameters are dependent, i.e. a large deviation in one parameter can be compensated by a corresponding deviation of the other. This can be overcome by transforming the lg*a*-parameter to an independent constant. This is accomplished by transforming the model to

$$\lg N = \lg \tilde{a} - b\left(\lg S - \overline{\lg S}\right) + \varepsilon \tag{2}$$

where $\overline{\lg S}$ is the average of the experimental logarithmic stresses and the new independent parameter $\lg \tilde{a}$ is

$$\lg \tilde{a} = \lg a - b \overline{\lg S}$$

The results are compared in table 5, and the corresponding Wöhler curves are plotted in figure 1.

Table 5. The estimated parameters in the Wöhler curve as they have been calculated and reported by each laboratory

	A	B	$\lg a$	b	$\overline{\lg S}$	$\lg \tilde{a}$
lab 1	2.9429	0.065	45.3	15.4	2.59	5.41
lab 2	-	-	33.2	10.7	2.60	5.38
lab 3	-	-	30.4	9.6	2.58	5.63
lab 4	2.838	0.0365	77.8	27.4	2.59	6.83
lab 5	-	-	-	-	-	-
lab 6	-	-	-	-	-	-

It can be seen that considerable differences appear between laboratories and it is apparent that the one laboratory that makes the estimation in the opposite direction and includes the run outs is the extreme case.

An approximate statistical test (see appendix) shows a significant laboratory effect, i.e. the differences in Wöhler curves cannot be explained by material scatter only.

In order to investigate if the laboratory effect is solely caused by the modelling uncertainty we estimated new parameters from the raw data with a common algorithm. We then chose to use all failed specimens and made the minimization in the logarithmic life direction. The results are presented in table 6 and the corresponding curves are shown in figure 2.

Table 6. The estimated parameters in the Wöhler curve using common procedure

	$\lg a$	b	$\overline{\lg S}$	$\lg \tilde{a}$
lab 1	45.3	15.4	2.59	5.45
lab 2	45.8	15.5	2.60	5.51
lab 3	29.7	9.4	2.58	5.46
lab 4	48.5	16.6	2.59	5.49
lab 5	43.4	14.6	2.60	5.44
lab 6	37.5	12.4	2.60	5.28

Comparing tables 5 and 6 shows clear differences for laboratories 2 and 4. The reason for the laboratory 2 deviation is that they only used the two upper levels for their regression, but we here include the one failure at the lower level. The laboratory 4 deviation depends on two things. At first they have included the run outs in the regression. Secondly they have made the regression in the opposite direction.

The results in table 6 can be used in a formal statistical significance test. The result of such a test shows **no evidence for a laboratory effect** (see appendix).

Table 7. Experimental results, number of cycles to failure

Level	Test	Nom S_{max}	Nom S_{min}	Lab 1	Lab 2	Lab 3 Act Stress* S_{max}	Lab 3 Act Stress* S_{min}	Lab 3	Lab 3 adj.**	Lab 4	Lab 5	Lab 6***
1	1	400	40	run-out	1473253	397.61	40.00	607997	570219	run out	1483571	run out
	2	400	40	724017	run-out	397.63	40.11	779516	729311	run-out	261716	-
	3	400	40	run out	run out	397.74	40.23	538518	503699	2568575	run out	-
	4	400	40	1307156	run out	397.97	40.58	707232	659369	568310	run out	-
2	5	430	43	232775	261641	424.01	44.20	384613	321061	311015	174876	191655
	6	430	43	160231	179983	425.65	41.31	243643	227985	270718	180658	125506
	7	430	43	375570	366804	427.15	43.23	125877	116554	260942	122797	245712
	8	430	43	279472	266422	427.18	43.62	273603	251060	199419	429765	227409
3	9	460	46	110582	190827	457.61	46.91	182221	168704	115867	109815	74214
	10	460	46	113672	103165	458.03	46.15	151032	143746	101727	127295	75586
	11	460	46	93687	88472	459.15	46.18	177979	173760	111429	96859	102128
	12	460	46	113556	146211	460.18	46.15	154493	154601	112315	165506	105819
		460	46	-								123250

* Lab 3 measured the dynamic stress during the test and reported it

** The results of Lab 3 were adjusted to give the same S_{max} and S_{min} as the other

*** Lab 6 did not test more than one specimen on the lowest level instead five specimens were tested on the highest level

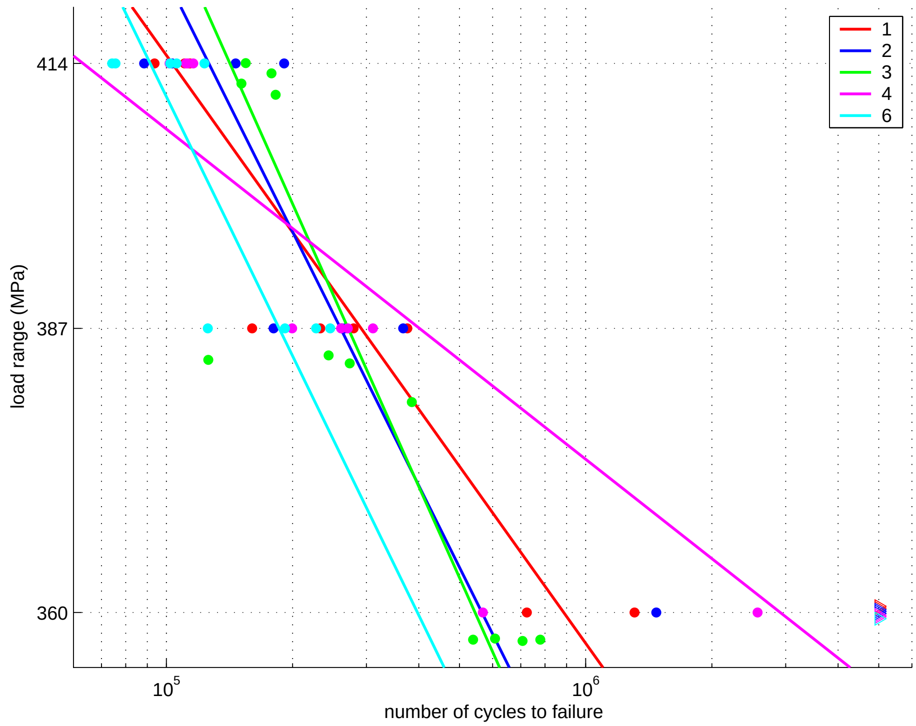


Figure 1. All experimental results and estimated Wöhler curves from the different laboratories.

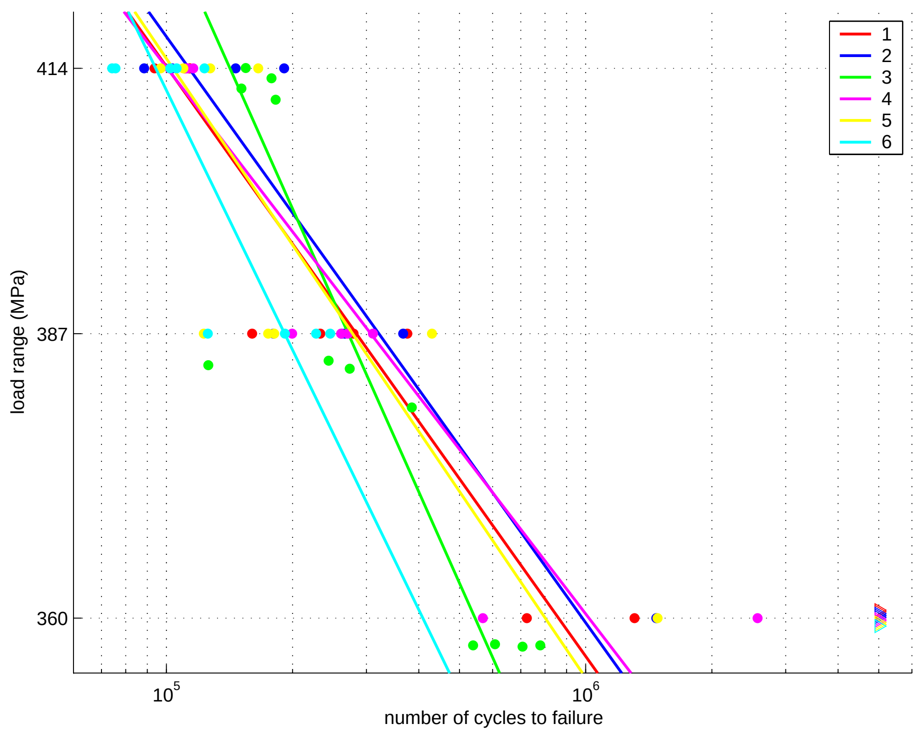


Figure 2. All experimental results and estimated Wöhler curves using the common procedure.

DISCUSSION

Experimental results

Fatigue tests are difficult to perform and in this study a few specimens have been tested in an incorrect way. The results of these tests have not been used in the statistical analysis. Some laboratories detected a temperature rise (maximum 75°C) in the test specimens when testing on the 460 MPa level. The temperature rise was considered to have no significant effect on the results and was therefore neglected.

Most laboratories have performed estimations of the Wöhler curve parameters. Visual comparison of their estimated curves suggests differences and a statistical test verifies the conclusion: There is a statistically significant laboratory effect.

A closer study of each participant's procedure for determining the Wöhler curve shows that the differences seem to be caused by different modelling of the curve. The linear regression is made in different directions, i.e. the error minimization is made either in the direction of logarithmic life or the logarithmic load. Another difference is the interpretation of the results on the lowest load level. Most participants exclude the run out results, one participant includes them and one excludes all results on this lower level.

Next we tried to find out if there remain any laboratory differences after excluding the model interpretation effects. This was accomplished in two ways, namely firstly by direct comparison of the obtained experimental fatigue lives and secondly by using the same estimating procedure on all data sets. The first comparison was done on the two higher load levels. For these, no statistically significant differences were found. The second comparison, which includes the failures on the lowest level, verifies the result.

The conclusion is that no systematic errors in measurements have been detected, but different modelling techniques give significant differences in the results. This important observation gives rise to the question: Is there any modelling procedure that is or can be agreed upon in the fatigue society? If not, it is of utmost importance that the modelling procedure is clearly defined in the report. It is very important for the laboratories' customers to be aware of this fact and, when ordering a test, to ask for a preferred modelling procedure as well as being aware of the modelling procedure used by the laboratory when using fatigue data in design.

One factor, which is problematic to deal with, is the order the test specimens are tested in. The most correct way would be to set up the test program where both the specimens and the different load levels were randomised. This is usually not practical in a normal laboratory situation.

One of the participants found that the original specimens were bended to different degrees. This observation suggests that the original steel plates had different bending properties, which could affect the results. This effect should not influence the lab to lab variation since the specimens were randomised between laboratories, but it is still interesting to evaluate. In order to see the plate influence we estimated a common Wöhler curve from all experimental results and calculated the residuals from this estimation. The residuals are the absolute values of the deviations in log lives from the estimated curve. The result is illustrated in figure 3 indicating that plate 5 has shorter lives than the other. This is in accordance with the bendiness measurements.

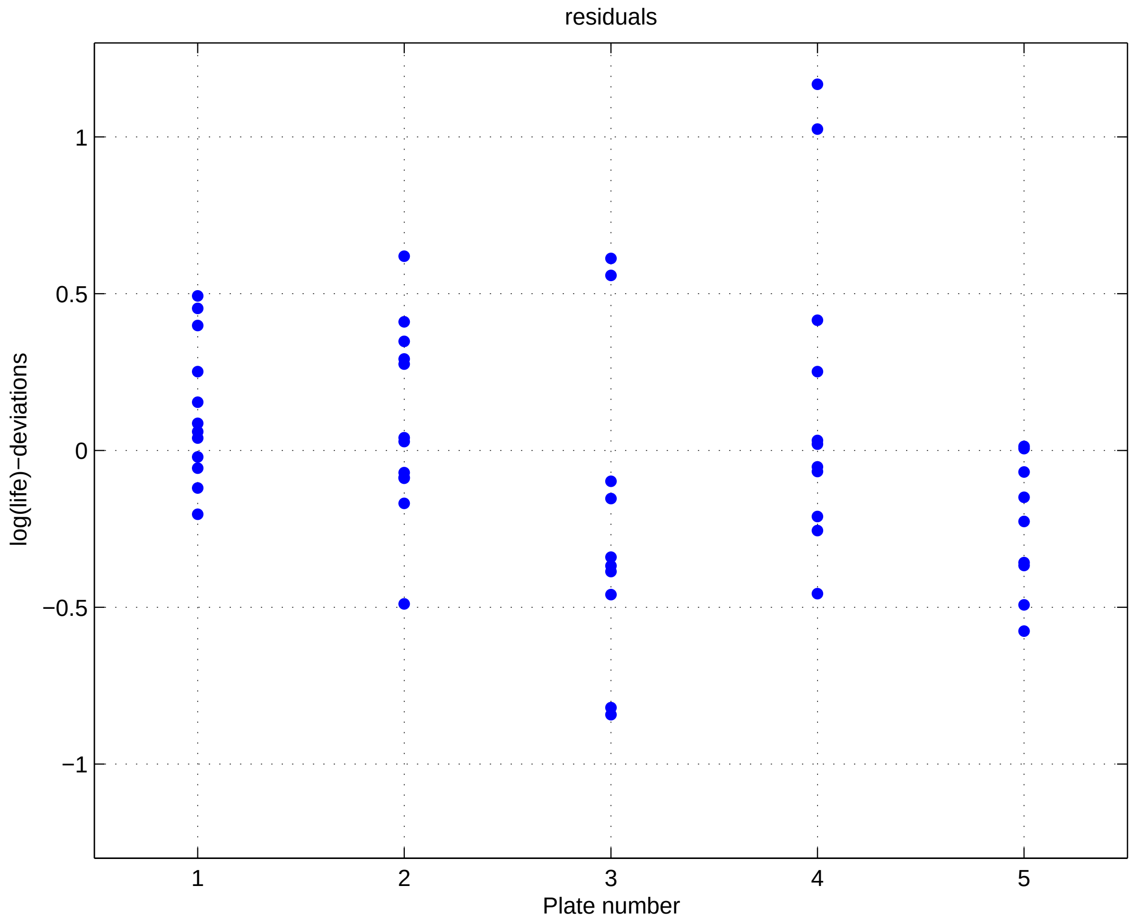


Figure 3. Deviations from a common Wöhler curve plotted against the plate number.

Uncertainty evaluation

All laboratories have made some considerations regarding the uncertainties in measurement. However, none of them have evaluated uncertainties for the resulting Wöhler parameters, but only for the applied stress.

Only one participant has used the method recommended by the ISO guide, GUM. This is surprising since the GUM guide has been recommended by European accreditation authorities for several years.

Among the uncertainty sources that have been identified by the laboratories only load cell measurement uncertainties and dimension measurement uncertainties have been taken into account. Important sources like misalignment and load control have been identified by some participants but not included in the evaluation of the stress uncertainty. Apparently only calibrated devices are considered for the overall uncertainty and other sources, more difficult to evaluate, have been excluded. No motivation for these exclusions can be found in the reports.

One participant has rejected the uncertainty evaluation with reference to the large scatter in fatigue lives. Our overall conclusion from the laboratory comparisons that there are no detectable systematic effects may be seen as verification of this rejection, but it is questionable if this was an obvious result beforehand. In addition, for instance, uncertainties due to misalignment are not obviously negligible compared to the material scatter and should be considered in an uncertainty analysis.

CONCLUSIONS

Laboratories were invited to participate in the inter-laboratory comparison through an open call on Internet and in the UTMIS network (with in total 31 member organisations) and 8 were specially invited with the aim of reaching at least ten participants. Seven laboratories agreed to participate and six of them produced results.

The raw data show no laboratory effects but differences in modelling (due to run-outs, regression method) result in significant laboratory differences. Modelling effects are difficult to detect for a single laboratory and the observed differences emphasize the importance of careful definition of the models, both beforehand and in the report.

How to define, calculate, and interpret measurement uncertainty and to use it in Wöhler-curve determination is poorly understood among the participants, in spite of the fact that they consist of a group with significant experience of fatigue testing. An important overall tendency is that the laboratories only include uncertainty sources that are easily obtained, e.g. from calibrated gauges.

RECOMMENDATIONS

It is important to include original data in reports to allow for re-evaluation in case of ambiguities about modelling or calculation errors.

It is also important for the laboratory to understand that since its customers may not be experienced in fatigue testing, fatigue test results must be reported with sufficient and clear information to make it possible for the customer to take decisions or to use the results in design situations.

To make uncertainty calculations easier to perform, a checklist like the one in appendix 2 could be used.

Instructions for how to evaluate the different uncertainties for each specific application should be created, in particular for sources that are not under full control.

CONTINUATION

The project has been successful and there are several possibilities for continuations. One is to isolate specimen bending from machine misalignment and thereby perform a more thorough uncertainty analysis of the fatigue test results. It would also be of interest to further develop methods to handle the other uncertainty sources.

Another possible continuation of this project is to design a course for fatigue test laboratories in uncertainty analysis of fatigue properties.

The possibility for publication of the results will be judged in the near future as additional information arrives from the different testing laboratories. Either this will give a sufficient basis for publication or the proposed continuations are needed in addition.

ACKNOWLEDGEMENT

This project was financially supported by NORDTEST and the participating organisations funded their own participation. The discussions with the participants have been very valuable and have given important input to the report.

REFERENCES

[1] ISO/IEC 17025:1999 General Requirements for the competence of testing and calibration laboratories, ISO Geneva 1999.

[2] ASTM E-466-96 Standard Practice for Conducting Force Controlled Constant Amplitude Axial Fatigue Tests of Metallic Meterilas.

[3] ISO5725-2 :1994(E) Accurency (Trueness and precision) of measurement and result. Part 2 Basic method for determination of repeatability and reproducibility of standard measurement method.

[4] BIPM, IEC, IFCC, ISOIUPAC, IUPAP, OIML, Guide to the expression of uncertainty in measurement (GUM), ISBN 92-67-10188-9, ISO Geneva 1993 (corrected and reprinted 1995).

APPENDIX 1
Formal test on the parameters

When using linear regression on the statistical model (2) one obtains estimates of the parameters $\lg\tilde{a}$ and b. In addition one gets an estimate of the variance of the random component ε. All these three estimates are independent, and based on this fact we can construct significance tests for the hypothesis that the parameters are the same for all laboratories. The test for the $\lg\tilde{a}$ parameter is constructed as follows: If the null hypothesis is true the variance of the estimate can be calculated in two ways. The first estimate is simply the squared sample standard deviation of the six estimates,

$$\hat{\sigma}^2_{(1)}\left(\lg\hat{\tilde{a}}\right)=\frac{1}{5}\sum_{i=1}^{6}\left(\lg\hat{\tilde{a}}_i-\overline{\lg\hat{\tilde{a}}}\right)^2,$$

where $\lg\hat{\tilde{a}}_i$ is the estimate from laboratory i, and $\overline{\lg\hat{\tilde{a}}_i}$ is the average of the six estimates. In this case this property can be calculated from the last column of table 5, giving

$$\hat{\sigma}^2_{(1)}\left(\lg\hat{\tilde{a}}\right)=0.0067 \tag{3}$$

The second estimate is based on a weighted average of the individual $\lg\tilde{a}_i$ -estimate variances,

$$\hat{\sigma}^2_{(2)}\left(\lg\hat{\tilde{a}}\right)=\frac{\sum_{i=1}^{6}(n_i-2)\frac{s_i^2}{n_i}}{\sum_{i=1}^{6}(n_i-2)}$$

where s_i is the sample standard deviation from each regression, and n_i is the number of tests in each regression. This property is estimated at

$$\hat{\sigma}^2_{(2)}\left(\lg\hat{\tilde{a}}\right)=0.0027$$

Since the two variance estimates are independent one can perform an *F*-test:

$$f=\frac{\sigma^2_{(1)}\left(\lg\hat{\tilde{a}}\right)}{\sigma^2_{(2)}\left(\lg\hat{\tilde{a}}\right)}\sim F_{\nu_1,\nu_2}$$

where ν_1,ν_2 are the numbers of degrees of freedom for the two estimates. For the nominator estimate ν_1 is equal to the number of laboratory Wöhler curves minus one and for the denominator estimate ν_2 is equal to the total number of experiments minus the total number of estimated parameters.

If the null hypothesis is false, i.e. if we have a laboratory effect, then the first estimate should be larger than the second, which means that the null hypothesis is rejected for large values of *f*. Here we obtain,

$$f=\frac{0.0067}{0.0027}=2.48\sim F_{5,47}$$

and from the *F*-distribution we find

$$P(f>2.48)=5.7\%$$

which means that the null hypothesis cannot be rejected on a 5 % level.

The corresponding formulae for the *b*-estimate is

$$\hat{\sigma}^2_{(1)}\left(\hat{b}\right)=\frac{1}{5}\sum_{i=1}^{6}\left(\hat{b}_i-\overline{\hat{b}}\right)^2,$$

where $\hat{b}_i$ is the estimate from laboratory *i*, and $\overline{\hat{b}}$ is the average of the six estimates,

$$\hat{\sigma}^2_{(2)}\left(\hat{b}\right)=\frac{\displaystyle\sum_{i=1}^{6}\frac{(n_i-2)s_i^2}{\sum_j\left(\lg S_{ij}-\overline{\lg S_{i.}}\right)^2}}{\displaystyle\sum_{i=1}^{6}(n_i-2)}$$

where S_{ij} is the stress level in the *j*-th trial at laboratory *i,* and $\overline{\lg S_{i.}}$ is the average of the logarithmic stresses at laboratory *i*. The two variance estimates are independent and one can perform the *F*-test:

$$f = \frac{\sigma^2_{(1)}\left(\hat{b}\right)}{\sigma^2_{(2)}\left(\hat{b}\right)} \sim F_{\nu_1,\nu_2}$$

where ν_1, ν_2 are the numbers of degrees of freedom for the two estimates. If the null hypothesis is false, i.e. if we have a laboratory effect, then the first estimate should be larger than the second, which means that the null hypothesis is rejected for large values of *f*.

For the exponent we obtain the following values:

$$\hat{\sigma}^2_{(1)}\left(\hat{b}\right) = 7.01 \qquad \hat{\sigma}^2_{(2)}\left(\hat{b}\right) = 6.43 \qquad f = \frac{7.01}{6.43} = 1.09 \sim F_{5,47}$$

$$P\left(f > 1.09\right) = 38\%$$

There is no evidence that the exponents differ between laboratories.

In case of the originally estimated Wöhler curves the formal test is more difficult to evaluate, since the estimation procedure not is common. However an approximate test can be done, using the standard deviations of the laboratory parameter estimates and comparing with the denominator variances used above.

$$f^{(orig)}_{\lg\hat{\alpha}} = \frac{0.473}{0.0027} = 175 \sim F_{3,47} \qquad P\left(f^{(orig)}_{\lg\hat{\alpha}} > 175\right) = 0\%$$

where the three number of degrees of freedom for the nominator is because only four laboratories have reported any estimates. Here we see a significant difference between laboratories.

$$f^{(orig)}_{\hat{b}} = \frac{66.4}{6.43} = 10.3 \sim F_{3,47} \qquad P\left(f^{(orig)}_{\hat{b}} > 10.3\right) = 0.00002\%$$

Also the difference in exponents is significant.

Another test based directly on the experimental results

The formal test above is based on the estimated parameters. Another way to perform the test is to study the experimental lives on different levels under the assumption that the variance is the same on these levels. In our case the lowest level contains survivors which are difficult to take into account and also can be assumed to have a larger variance. Therefore only the two higher levels are investigated. A linear model is used

$$\lg N\left(S_{ijk}\right) = \gamma_k + \eta_j + \varepsilon_{ijk} \quad ,$$

where $\lg N\left(S_{ijk}\right)$ is the *i*-th experimental result for laboratory *j* on load level *k*, γ_k is the mean result on the *k*-th level, η_j is the effect of the *j*-th laboratory, and ε_{ijk} is the same random error as in (1). Using such a linear model makes it possible to perform a standard two-way analysis of variance.

One complication is that one of the laboratories has performed the test on load levels that deviates from the standard levels. Therefore we have adjusted these results to the standard levels by using the Wöhler curve obtained by that particular laboratory. Example: experiment no. 5, Actual load range is $S_{actual} = 424.01 - 44.20 = 379.81$. The Wöhler curve for this particular laboratory is

$$\lg(N) = 29.7 - 9.4 \lg S$$

For the actual load range we calculate the obtained residual

$$e = \lg(384513) - 29.7 + 9.4 \lg(379.81) = 0.13$$

The adjusted life for specimen no. 5 is then

$$\lg(N_{adj}) = 29.7 - 9.4 \lg(387) + 0.13 = 5.51, \qquad N_{adj} = 10^{5.51} \approx 320000$$

The result from the analysis of variance is seen in figure 4 where the deviations of the logarithmic lives from the overall load level mean, $\lg N(S_{ijk}) - \overline{\lg N(S_{ijk})}$, are illustrated in a box plot.

The formal standard F-test shows **no significant laboratory influence** (significance level 28 %), see the computer output ANOVA table below, where 'Columns' represent laboratories and 'Rows' represents load levels.

'Source'	'SS'	'df'	'MS'	'F'	'Prob>F'
'Columns'	[0.5856]	[5]	[0.1171]	[1.3256]	[0.2755]
'Rows'	[5.2045]	[1]	[5.2045]	[58.9087]	[4.3605e-009]
'Interaction'	[0.4710]	[5]	[0.0942]	[1.0662]	[0.3951]
'Error'	[3.1806]	[36]	[0.0883]		
'Total'	[9.4417]	[47]			

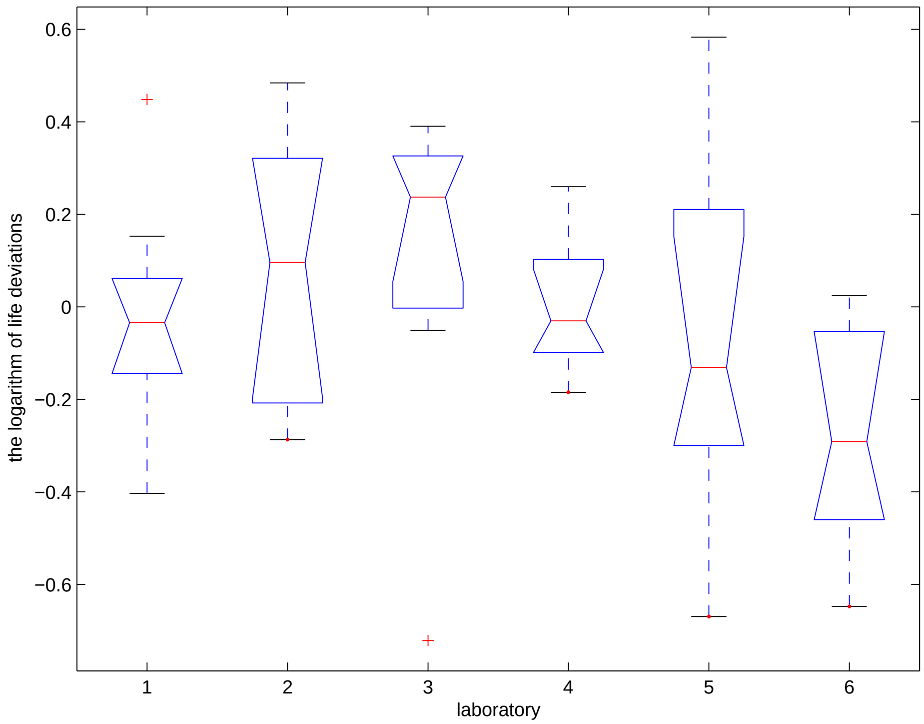


Figure 4. Box plot of the deviations from mean estimates. The different lines of the box plot represent (from the bottom to the top): the lowest line is the lowest result, the second line is the 25% percentile, the middle line is the median, the second line from the top is the 75% percentile and the maximum value is represented by the top line. Out-liers are marked with +.

APPENDIX 2
Checklist to be used when performing measurement uncertainty calculations.

The following sources of uncertainty should be considered when performing uncertainty calculations.

Influencing factor	**Comments**
Load control	The way the load is controlled may affect the applied stresses in the test specimens, e.g. a high frequency can lead to inaccurate maximum load
Bending stress due to mis-alignment of the test machine	
Bending stresses due to incorrectly manufactured test specimens	This is a specially important factor when testing welded specimens
Uneven stress field due to incorrectly mounted test specimens	This could be avoided by using special fixtures when mounting the test specimens
Dimensions	
Lab. temperature	Usually not influencing the result
Lab. humidity	Usually not influencing the result
Specimen temperature	There are situations (frequency and material dependent) where the temperature in the test specimen may rise enough to influence the fatigue strength
Corrosion	The test specimens could be affected by corrosion due to e.g. bad storage
Modelling	
Frequency	See above concerning specimen temperature and load control.
Failure criterion	Usually not influencing the result for test specimens, but could influence when larger structures are tested.
Scratches	If the specimens are not properly packaged, they may be damaged during transport

The following formula could be used.

Influencing factor	**Way of treating the influencing factor (neglect, mention in report or take into account)**	**Comments**
Load control		
Bending stress due to misalignment of the test machine		
Bending stresses due to incorrectly manufactured test specimens		
Uneven stress field due to incorrectly mounted test specimens		
Dimensions		
Lab. temperature		
Lab. humidity		
Specimen temperature		
Corrosion		
Modelling		
Frequency		
Failure criterion		
Scratches		

TECHNICAL REPORTS FROM EXPERT GROUP QUALITY AND METROLOGY

Notice: Only technical reports with a bold number on the left leaf of the page can be ordered free of charge from the Nordtest secretariat. Others have to be ordered from the publishing organisation or institute. Information for ordering those reports can be obtained from Nordtest secretariat and Nordtest Web-site.

403 Holmgren, M., Observing validation, uncertainty determination and traceability in developing Nordtest test methods. Espoo 1998. Nordtest, NT Techn Report 403. 12 p. NT Project No. 1277-96.

418 Views about ISO/IEC DIS 17025 - General requirements for the competence of testing and calibration laboratories. Espoo 1999. Nordtest, NT Techn Report 418. 87 p. NT Project No. 1378-98.

419 Virtanen, V., Principles for measuring customers satisfaction in testing laboratories. Espoo 1999. Nordtest, NT Techn Report 419. 27 p. NT Project No. 1379-98.

420 Vahlman, T., Tormonen, K., Kinnunen, V., Jormanainen, P. & Tolvanen, M., One-site calibration of the continuous gas emission measurement methods at the power plant. Espoo 1999. Nordtest, NT Techn Report 420. 18 p. NT Project No. 1380-98.

421 Nilsson, A. & Nilsson, G., Ordering and reporting of measurement and testing assignments. Espoo 1999. Nordtest, NT Techn Report 421. 7 p. NT Project No. 1449-99.

430 Rasmussen, S.N., Tools for the test laboratory to implement measurement uncertainty budgets. Espoo 1999. Nordtest, NT Techn Report 430. 73 p. NT Project No. 1411-98.

431 Arnold, M., Rooound robin test of olfactometry. Espoo 1999. Nordtest, NT Techn Report 431. 13 p. NT Project No. 1450-99.

429 Welinder, J., Jensen, R., Mattiasson, K. & Taastrup, P., Immunity testing of integrating instruments. Espoo 1999. Nordtest, NT Techn Report 429. 29 p. NT Project No. 1372-97.

443 Guttulsrød, G.F, Nordic interlaboratory comparison measurements 1998. Espoo 1999. Nordtest, NT Techn Report 443. 232 p. (in Dan/Nor/Swed/Engl) NT Project No. 1420-98.

452 Gelvan, S., A model for optimisation including profiency testing in the chemical laboratories. Espoo 2000. Nordtest, NT Techn Report 452. 15 p. NT Project No. 1421-98.

501 Lau, P., Study to characterize thermal convection effects in water. Espoo 2003. Nordtest, NT Techn Report 501. 15 p. NT Project No. 1543-01.

503 Erikoinen, O., Kiiskinen, J. & Pajari, M., Interlaboratory comparison tests of reinforcing steel roducts. Espoo 2003. Nordtest, NT Techn Report 503. 31 p. NT Project No. 1446-99.

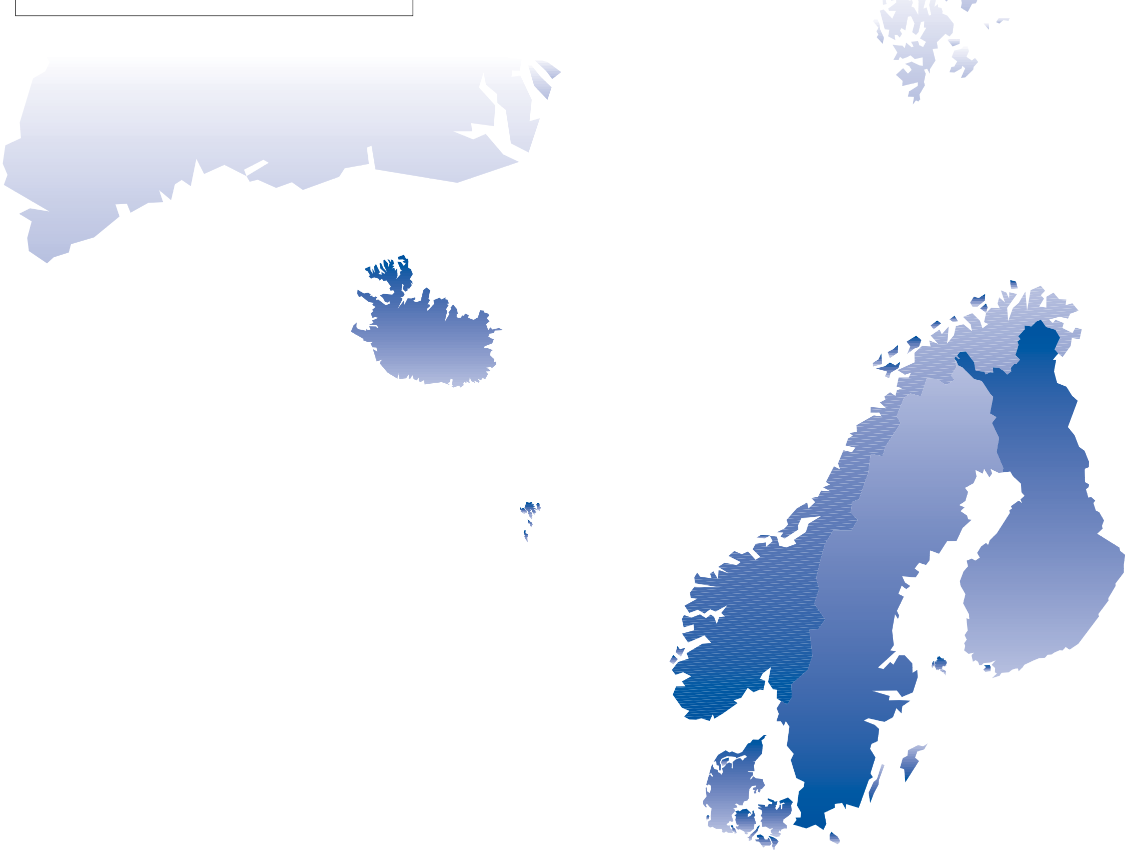
512 Merryl, J., Estimation of assigned values and their uncertainties for use in interlaboratory comparisons. Espoo 2003. Nordtest, NT Techn Report 512. 57 p. NT Project No. 1496-00.

513 Hovind, H., Severinsen, G. & Settergren-Sørensen, P., Nordic standard interface for transfer of data and graphics between proficiency test webs and statistical software. Espoo 2003. Nordtest, NT Techn Report 513. 72 p. NT Project No. 1542-01.

514 Petersen, L., Frølund, H. & Lorentzen, E., Qualification of personnel in laboratories, inspection and certification bodies. Knowledge management in laboratories. Espoo 2003. Nordtest, NT Techn Report 514. 29 p. Appendix 1: 514_A1-Knowledge management (power point presentation) 10 slides. NT Project No. 1564-01.

533 Svensson, T., Holmgren, M., Johansson, K. & Johnson, E., Inter-laboratory comparison of fatigue test with evaluation of the participating laboratories calculations of measurement uncertainty. Espoo 2003. Nordtest, NT Techn Report 533 20 p. NT Project No. 1591-02.

NORDTEST
TECHNICAL REPORT 533



Nordtest, founded in 1973, is an institution under the Nordic Council of Ministers and acts as a joint Nordic body in the field of conformity assessment. The emphasis is on the development of Nordic test methods and on Nordic co-operation concerning conformity assessment. The main task is to take part in the development of international pre-normative activity. Nordtest is yearly funding projects in its field of activity.

Nordtest endeavours to

- promote viable industrial development and industrial competitiveness, remove technical barriers to trade and promote the concept “Approved Once Accepted Everywhere” in the conformity assessment area
- work for health, safety, environment in methods and standards
- promote Nordic interests in an international context and Nordic participation in European co-operation
- finance joint research in conformity assessment and the development and implementation of test methods
- promote the use of the results of its work in the development of techniques and products, for technology transfer, in setting up standards and rules and in the implementation of these
- co-ordinate and promote Nordic co-operation in conformity assessment
- contribute to the Nordic knowledge market in the field of conformity assessment and to further development of competence among people working in the field

Nordic
COUNCIL OF MINISTERS