INTERNATIONAL SEED TESTING ASSOCIATION (ISTA)
Secretariat, Zürichstrasse 50, 8303 Bassersdorf, CH-Switzerland
Phone: +41-44-838 60 00, Fax: +41-44-838 60 01, Email: ista.office@ista.ch, http://www.seedtest.org

Tolerances in ISTA Moisture Testing
A discussion paper

1. Introduction and background
One of the objectives of ISTA is to improve the tolerances in the ISTA Rules. This paper describes the present situation of the tolerances in chapter 9, Determination of Moisture Content, of the ISTA Rules.

This information will be used for improving tolerances already present in the Rules, and for developing new tolerances for moisture test situations not yet covered.

Relevant parts from the ISTA Rules and the ISTA Moisture Handbook are in blue and in letter type Times New Roman.

2. Tolerances for replicates in one test
2.1 General – oven
The new ISTA Rules are as follows since the introduction per January 1st 2008:

9.1.6.2 Tolerances
Report the result as the arithmetic mean of the duplicate determinations carried out on a sample if the difference between the two determinations does not exceed the 0.2% tolerance. For species that need to be cut (Table 9.A.2) it has been found impossible to meet a 0.2% tolerance and tolerances ranging from 0.3 to 2.5% are prescribed. These are related to seed size and initial moisture content. (Table 9.B)

To use Table 9B, in column 1, select the relevant now depending on seed size. Then select the correct column (2, 3 or 4) depending on the initial moisture content of the sample.

If the results of the duplicate determinations are out of tolerance, repeat the test beginning at 9.1.5.2. For repeated tests, report the result of the second test if its replicates are in tolerance. If the replicates of the second determination are out of tolerance as well, check if the averages of the two tests are in tolerance (0.2%). If so, report this average. If replicates of both tests are out of tolerance and the average results of the repeat tests are out of tolerance discard the results, check the equipment, the laboratory procedure and start again.

Table 9.B.
Tolerance levels for differences between the two duplicate determinations of moisture content of tree and shrub seeds (significance level not defined).

Seed size class	Average initial moisture content		
	<12%	12 to 25%	>25%
1	2	3	4
Small: TSW <200 g	0.3%	0.5%	0.5%
Large: TSW >=200 g	0.4%	0.8%	2.5%

(Source: F.T. Bonner: Tolerance limits in measurement of tree seed moisture. Seed Science & Technology **12**, 789-794, 1984 [Table 3]).

No evidence is available on how the present tolerances for differences between two replicates in one test (0.2%) have been entered in the Rules. These tolerances have been inserted in the Rules before 1970. It is most probable that the 0.2% is based on experience, not on statistics.
In the MOI report of the period 1977-1980 (SST 9 (1981) 229-254) it was concluded that the existing tolerances were too narrow for tree seeds (p. 233-234). Later tolerances for tree and shrub seeds were changed.

The choice for value of 0.2% tolerance in moisture determination is pertinent because very close to the repeatability indicated in ISO 712 (0.15% for cereals) and in ISO 665 (0.2 % for sunflower and rape).
In ISO 665, it is clearly indicated that the repeatability value has been estimated after three comparative testing organised in 1986 and 1987 among 15 laboratories that analysed two replicates.

Precision = repeatability (within lab) + reproducibility (between labs).
Jean-Louis (email 2006-12-22) made some calculations based on the dataset of the comparative testing (whole seeds, both at 103°C and 130°C). At 99.8% confidence level (6 sigma) the precision would be 0.5%; at 95% confidence level it would be 0.26%. For MOI to decide what to do: what confidence level, in relation to the feeling about the percentages mentioned.
Again, these percentages are based on the comparative testing, which means: whole seeds only!

2.2. General-meters

9.2.2.6 Tolerances

The result is the arithmetic mean of the duplicate determination, if the difference between the two determinations does not exceed 0.2%

If the results of the duplicate determinations are out of tolerance, repeat the test. For repeated tests, report the result of the second test if its replicates are in tolerance. If the replicates of the second measurement are out of tolerance as well, check whether the averages of the two tests are in tolerance (0.2% or Table 9B). If so, report this average. If replicates of both tests are out of tolerance and the average results of the repeat tests are out of tolerance, discard the results, check the equipment, the laboratory procedure, and start again.

The reported result is rounded to one decimal place.

2.3. Effects of temperature and grinding

2.3.1. Effects of temperature in the oven

In the oven test, temperature may be of influence for the precision of the test. This could be deducted from the results of the comparative testing carried out in 2005.

2.3.2. Effects of grinding (Jean-Louis Laffont)

Three sources of data were considered to assess the differences in moisture precision for Whole vs Ground:
. Proficiency test data of Trifolium inc (Whole) and Triticum aest (Ground).
. Ring test data performed within the European Grain network (all Ground).
. Comparative testing data (all Whole).
These data have been analyzed using a linear mixed-effects model; repeatability and reproducibility were then estimated after fitting the models. The details of these analyses are given in annexes 2.3.2.1, 2.3.2.2 and 2.3.2.3.
The table below summarizes the results:

			Repeatability	Reproducibility
Proficiency test data	**Whole**		0.0459	0.4838
	Ground		0.1487	0.5379
Ring test data	**Ground**	**Wheat**	0.0056	0.0302
		Barley	0.0057	0.0280
Comparative testing data	**Whole**	**103°C - 17h**	0.0063	0.0235
		130°C - 17h	0.0095	0.0284

The Proficiency Test (PT) results, which has both Whole and Ground data exhibits larger repeatability for Ground. The magnitude of the repeatability and reproducibility for this test is very different from the magnitude of these quantities for the Ring test and the Comparative testing, which both produce similar estimates.
There appears to be a difference between whole and ground.

Grinding is a supplementary step, adding error to the result.
This is reflected in the statistics: repeatability (differences between reps in one lab) is three times worse for ground seeds compared to whole seeds.
Reproducibility (differences/spread between the results of different labs for one sample) is the same for ground and whole seeds. Not having a difference in reproducibility may mean that the factor grinding is far less important than other steps in the method, like oven type, temperature, duration, balance. This has to be further explored.

As a consequence of the differences in repeatability for whole and ground, the tolerances between replicates in one lab would be three times larger for whole compared to ground seeds.
In addition: tolerances for comparing two results of one sample in different labs would be more or less the same for whole and ground.
However, as these statistics are based on results of which the retests have been excluded, no tolerances can be calculated yet.

The better repeatability for whole compared to ground may be a general feature.
This is assuming that after 17hrs103°C all water will have had the opportunity to pass through the seed coat.
From the comparative testing of 2005 it appears that neither oil content, nor thousand seed weight is related to any of the methods. This suggests that seed coat may be the most important factor for the speed of loss of water from seeds. We know it plays a role in the short duration test at 130°C, f.e. for grasses and clovers. It probably will not be a problem for the long duration test.

The level of repeatability and reproducibility are quite different depending on the test rounds. The PT results exhibit very different repeatability and reproducibility magnitudes compared to the comparative testing. From the difference between the ring test data and the comparative testing analysis, it can be concluded that whole and ground are equivalent. But this was after removing some large outside values from the ring test, which can be an indication of potential problems with ground seeds.

Furthermore the PT data are of two different species. This may have had an influence on the outcome as well.

Therefore we cannot calculate a multiplier from whole to ground, f.e. based on the results of the comparative testing.

3. Tolerances for comparisons of two tests – same sample/same lab
3.1. Oven-oven
No data available.

3.2. Oven-meter

9.2.2.8 Routine Checking of moisture meter and oven moisture content results

Table 9.D. should be used when checking moisture meters against oven results.

Table 9.D

Tolerance limits for differences between constant temperature oven moisture determinations and moisture determinations conducted using moisture meters

Chaffy Seeds		**Non-chaffy Seeds**	
Constant Temperature Oven Result (average of results)	**Tolerance**	**Constant Temperature Oven Result (average of results)**	**Tolerance**
Less than 10.9 %	0.5	Less than 11.3 %	0.4
11% - 12.9%	0.6	11.3% to13.7%	0.5
13% - 14.9%	0.7	13.8% to 16.2%	0.6
15% - 16.9%	0.8	16.3% to 18.0%	0.7
17.0% - 18.0%	0.9		

For check samples, a maximum of 5% may have a difference greater than the maximal permissible difference. If more than 5% of the samples have differences greater than this, a new calibration is required (see 9.2.1.).

Table 9D (comparison of moisture meter with oven) was calculated using the formulas in section 9.2.1.6.3.

3.3. Meter-meter

9.2.2.9 Checking of results from different moisture meters

Table 9.E. should be used when checking moisture meters against each other.

Table 9.E
Tolerance limits for differences between moisture determinations conducted using different moisture meters.

Chaffy Seeds		Non-chaffy Seeds	
Moisture content (average of two results)	**Tolerance**	**Moisture content (average of two results)**	**Tolerance**
Less than 10.5 %	1.0	Less than 10.7 %	0.8
10.5% to 11.4%	1.1	10.7% to 11.8%	0.9
11.5% to 12.4%	1.2	11.9% to 13.1%	1.0
12.5% to 13.4%	1.3	13.2% to 14.3%	1.1
13.5% to 14.4%	1.4	14.4% to 15.6%	1.2
14.5% to 15.4%	1.5	15.7% to16.8%	1.3
15.5% to 16.4%	1.6	16.9% to 18.0%	1.2
16.5% to 17.4%	1.7		
17.5% to 18.0%	1.8		

Table 9E (comparison results of two moisture meters) is an addition. It was derived as follows:
The ISTA Rules do not give tolerances for permissible differences between moisture content results obtained using different moisture meters on the same sample in the same or different laboratories. However, if the moisture meters are calibrated according to the ISTA Rules then theoretically the maximum permissible difference that can be obtained, if the meters being compared are within tolerance of the calibration limits, is twice the difference between the oven result (the reference method) and the Moisture meter result given in ISTA Rule 9.2.1.6.3 Tolerances.
For example if we have two meters that are calibrated according to ISTA and one moisture meter has the maximum positive permissible difference and the other has the maximum negative permissible difference then the total difference between the two meters is twice the permissible difference. This is the basis of Table 9.E, which should be used when comparing results of moisture determinations carried out on the same sample in the same or different laboratories using different moisture meters.

Secretariat, Zürichstrasse 50, 8303 Bassersdorf, CH-Switzerland
Phone: +41-44-838 60 00, Fax: +41-44-838 60 01, Email: ista.office@ista.ch, http://www.seedtest.org

4. Tolerances for comparisons of two tests – different sample/different labs

4.1. Oven-oven

No data available.

4.2. Oven-meter

No data available.

4.3. Meter-meter

No data available.

5. Validation of new species or new methods (from chapter 8 of the ISTA Moisture Handbook)

5.1. Oven

8.2. Method

The new basic reference method is 17 hours at 103˚C which is already considered as being a 'safe' method for species containing volatiles in the ISTA Rules (2006 Edition). The only aspect that needs to be tested is the necessity for grinding (8.2.1.). Acceptance of a shorter time than currently specified for the high temperature test at 130˚C may be possible after comparative testing (8.2.2.).

If a shorter test at 130˚C is accepted, the longer test at 103˚C is still valid. As a consequence, for species to which this applies the method used has to be specified on the Certificate.

8.2.1. Test necessity for grinding

The necessity for grinding depends on factors like seed size and seed coat permeability to water. Testing the effect of grinding is compulsory before a new species can be introduced into the ISTA Rules.
The basic reference method is 17 hours at 103˚C.
To test whether grinding is necessary, all samples of a species should be tested at the same time. If this is not possible, a minimum of one replicate per moisture content by grinding combination must be put together into one oven at one time.

The following guidelines should also be adhered to:

a. Each seed sample is to be processed and tested in two ways: whole and ground. If the seeds are so small that more than 55% passes a 2.0 mm round holed sieve, only fine grinding shall be applied.
b. Each seed sample to be tested should be at two, or more, moisture levels. These moisture levels should be within the normal range for the species and within the range that is routinely observed. One of the two moisture levels must be less than average, the other more than the average and both should be at least 5% moisture content apart.
c. At least two and up to four at a maximum, samples should be tested per moisture level.
d. Two replicates on each sample should be tested.
e. The test should be run by at least three laboratories, up to a maximum of six.

The minimum number of samples per laboratory: 2 (± grinding) x 2 (moisture levels) x 2 (samples) = 8 samples x 2 replicates = 16 containers.

The maximum number of samples per laboratory: 3 (± grinding/fine and coarse) x 2 (moisture levels) x 4 (samples) = 24 samples x 2 replicates = 48 containers.

The tolerance and the methodology to evaluate the test results is the same as the one described for acceptance of 130˚C, see 8.2.2. The dataset available so far has not given evidence of having a separate tolerance level for comparing results from ground and whole seeds (unpublished data from the moisture committee).

Evaluation of the test results for grinding:
A whole seed test is considered acceptable if the results do not differ significantly from the results obtained with ground seeds.
In order to define what is an acceptable difference, accuracy, repeatability and reproducibility are taken into account.

A tolerance value is used to identify all individual whole seeds results which deviate too much from the ground seeds mean per sample, and a maximum % of deviating results is used as a decision rule. The procedure is:

1. Compute the (laboratory x sample) means for the reference method (ground seeds) and the whole seeds method.
2. Compute the sample means for the reference method.
3. Compute the differences between the (laboratory x sample) means for the candidate method and the sample means for the reference method.
4. If 75% or more of these differences are within the tolerance range of ± 0.3%, accept the whole seeds method.

The level 75% is based on past experience, and is safe and practicable. It is still applicable when only three laboratories participate.
For an example with data and how the tolerance value is established see 8.2.2.

8.2.2. Test for acceptance of a shorter duration test at 130°C
This test is not compulsory; only in the case where a request for the short test is received, is it necessary to be carried out. Compare the method resulting from 8.2.1 (whole or ground) in a comparative test.
Test all samples at one time. If this is not possible, then a minimum of one replicate per sample must be put into one oven.

a. Each sample is to be processed and tested in one way, whole or ground, depending on the results from 8.2.1.
b. At least two moisture levels that are representative for the species and routinely observed must be used; the moisture levels must be less than average and more than average, and be at least 5% apart.
c. For five modalities: 1, 2, 3 and 4 hours at 130˚C, and 17 hours at 103˚C.
d. For each moisture level, at least two and up to four at a maximum, samples per modality.
e. With two replicates on each sample received.
f. At least by three laboratories, up to a maximum of six.

Evaluation of the test results for high temperature:
A shorter test at 130°C is considered acceptable if the results at this temperature do not differ significantly from the results obtained at 103°C for 17 h which are considered as the reference method.
In order to define what is an acceptable difference, accuracy, repeatability and reproducibility are taken into account.

A tolerance value is used to identify all individual results at 130°C which deviate too much from the 103°C for 17 h mean per sample, and a maximum % of deviating results is used as a decision rule. The procedure is:

1. Compute the (laboratory x sample) means for the reference method (103°C for 17 h) and the 130°C methods.
2. Compute the sample means for the reference method.
3. Compute the differences between the (laboratory x sample) means for the candidate method and the sample means for the reference method.
4. If 75% or more of these differences are within the tolerance range of ±0.3%, accept the corresponding 130°C method.
5. Select the 130°C method from the retained ones which have the shortest duration test.

The level of 75% is based on past experience, and is safe and practicable. It is still applicable when only three laboratories participate.

8.2.3. Example of evaluation of the test results

From three participating laboratories with two moisture levels (low and high) and three samples per modality (1, 2, 3): The table shows the evaluation of 130˚C for 1 hour against the reference method of 103˚C for 17 hours. The same methodology also applies for 2, 3 and 4 hours at 130˚C.

In the example, five differences (marked with an *) of 18 are out of the tolerance range. Consequently, 130˚C for 1 hour is not acceptable as a short test, because 72.2% is below 75%.

Sample	Moisture level	Laboratory	Reference results 17 h at 103˚C	Results 1 h at 130°C	Difference from the reference sample mean
1	low	A B C	**9.9** **10.0** **9.8**	10.0 9.9 9.6	0.10 0.00 -0.30
Reference 1 low sample mean		**9.90**			
2	low	A B C	**7.9** **8.1** **8.0**	8.1 8.6 8.4	0.10 **0.60*** **0.40***
Reference 2 low sample mean		**8.00**			
3	low	A B C	**8.9** **9.1** **8.5**	9.2 9.2 8.9	**0.37*** **0.37*** 0.07
Reference 3 low sample mean		**8.83**			
1	high	A	**12.8**	12.7	-0.23
		B	**12.9**	12.9	-0.03
		C	**13.1**	13.4	**0.47***

Reference 1 high sample mean		**12.93**			
2	high	A	**13.5**	13.6	-0.07
		B	**13.7**	13.7	0.03
		C	**13.8**	14.0	0.33
Reference 2 high sample mean		**13.67**			
3	high	A	**13.0**	13.3	0.03
		B	**13.3**	13.4	0.13
		C	**13.5**	13.5	0.23
Reference 3 high sample mean		**13.27**			

In this table we take the difference with the reference mean rather than the reference tested in the lab being compared. Doing this 83% would be in tolerance and the results would be OK to accept the quick method.
An alternative would be taking the result from the lab rather than an average we are comparing like with like. If it is averages then why not the average for the quick methods against the average for the reference? If we do this then only one of the differences is more than ±0.3 and again 83% would be in tolerance.
However, The reason why we are comparing to the sample mean across labs is because this mean is a better estimate of the true (unknown) value of the sample (3 data points instead of one, lab effect corrected).
In the table we have 6 samples, all 6 are supposed do have a different moisture value.
The samples 1 2 and 3 low are low, but different, they are not repeats taken from the same seed lot, they are taken from 3 seed lots with low moisture.
For each of the 6 samples; A B C samples prepared for the three labs are taken from the same seed lot, and are expected to have a similar moisture value.
We do not consider the comparison to the general mean as this will not take into account some eventual Lab x Sample effects which were considered when we established the 0.3% tolerance according to Miles general guidelines.

The tolerance value of 0.3% has been determined as described below:

8.3. Tolerance determination

The tolerance is computed using the data from the 2005 comparative test.

As described in the Miles Handbook of Tolerances and of Measures of Precision for Seed Testing (1963), first assess the magnitude of variation associated with the different sources of variation in moisture testing for the reference method (103°C for 17 h). This is done through fitting the following mixed-effects model using REML:

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + e_{ijk}$$

where:
y_{ijk} is the moisture observed for sample *i,* in laboratory *j,* and replicate *k*
$i = 1, \ldots, 20$ $j = 1, \ldots, 10$ $k = 1, 2$
μ is the intercept
α_i is the fixed effect of sample *i*
b_j is the random effect of laboratory *j*. The b_j are iid N(0, σ_L^2)
$(\alpha b)_{ij}$ is the random interaction effect between sample *i,* and laboratory *j*. The $(\alpha b)_{ij}$ are iid

N(0, σ^2_{SxL})

e_{ijk} are the residuals. They are iid N(0, σ^2).

From fitting this model, the variance component estimates are:

$\sigma^2_L = 0.01292$

$\sigma^2_{SxL} = 0.01059$

$\sigma^2 = 0.006274$

Then, the total variance of the mean of the two replicates is estimated by:

$$\sigma^2_{Total} = \sigma^2_L + \sigma^2_{SxL} + \frac{\sigma^2}{2} = 0.01292 + 0.01059 + \frac{0.006274}{2} = 0.026647$$

Assuming normal distribution of the mean of the two moisture replicates (a very reasonable assumption), then a 95% confidence is given by $\pm 1.96\sqrt{\sigma^2_{Total}} = 0.32$ which, rounded to the one decimal place leads to the tolerance 0.3%.

5.2. Moisture meters (calibration of moisture meters)

The calibration of moisture meters, as described in chapter 9.2.1 of the ISTA Rules, in fact is a validation procedure for moisture meters. Tolerances involved in this procedure can be found in paragraph 9.2.1.6.3:

9.2.1.6.3 Maximum permissible differences

A moisture meter is considered to be within calibration when z_i (the difference between y_x and the true value x_t) is lower than the following maximum permissible differences (Table 9.C):

Table 9.C.
Permissible differences from the true value

True value (reference method)	Maximal permissible difference	
	Non-chaffy seeds	**Chaffy seeds**
Less than 10.0%	: ± 0.4%	: ± 0.5%
10.0% or more	: ± 0.04 x moisture content	: ± 0.05 x moisture content

For the comparison the average result of the replicates after rounding to one decimal place must be used.

6. Discussion and conclusions

The present situation for tolerances in ISTA moisture testing can be summarized as in Table 6.1.

Table 6.1.
Tolerances in moisture testing.

Type of tolerance	Oven/Meter	Type of seeds	Tolerance
1a. Replicates in one test (same sample, one lab)	Oven	Agricultural seeds	0.2%
1b.	Oven	Tree & shrub seeds	Table 9B
1c.	Meter		0.2%
2a. Comparisons of routine tests (same sample, one lab)	Oven-Oven		N/A
2b.	Meter-Meter		Table 9E
2c.	Oven-Meter		Table 9D
3a. Comparisons of routine tests (different samples, same/different labs)	Oven-Oven		N/A
3b.	Meter-Meter		N/A
3c.	Oven-Meter		N/A
4a. Validation of new methods/species (different samples, different labs)	Oven		Max. 0.3% for 75% of results
4b.	Meter		Table 9C

For 4 of the 11 situations as described in Table 6.1 we don't have any tolerances.
The seven situations for which we have tolerances, only two are based on scientific evidence.
Situation 1b is based on experiments that were published in SST (see paragraph 2.1 of this paper for details).
Situation 4a is derived from the comparative testing MOI organized in 2005 (see paragraph 5.1.5. of this paper for an explanation).

The origin of the 0.2% in situation 1a is not known.
The tolerances of situation 1c are based on 1a.
Situations 2b, 2c and 4b are based on results from other organizations.

When we proposed to include Table 9C into the Rules (situation 4b.) barley was a chaffy seed and of the cereals only maize and wheat were non-chaffy. The purity definitions have now changed and barley is considered non-chaffy. This may be OK for many tests but for moisture testing it is not. Having a fused seed coat like barley makes it like the other chaffy seeds and causes increased variability in comparison to a non chaffy seed like wheat. It is suggested to change the wording because of the purity committees decision to make barley a non-chaffy seed.

It is suggested to make a Rules Change Proposal for 2009 for barley, and if relevant also for oats.

The ISTA Moisture Committee has discussed the possibility of using the results of ISTA Moisture content Proficiency Tests (PT) for the purpose of developing tolerances for seed moisture testing. However, the laboratories are requested to only report results when they are within tolerances. Using the PT results would thus result in a to narrow tolerance.

Results of recent comparative testing cannot be used for calculating tolerances between labs: as we didn't have any repeat samples per lab, the within lab tolerance has to be calculated from other sources.

The Committee has also discussed the possibility of using the results of other organization's proficiency tests results and/or use of other organization's tolerances (ISO, ICC, OIML, USDA, BIPEA-France, ….). It appears that results from comparative testing by these organizations are mostly not published in scientific magazines nor are available from these organizations.

It is suggested to gather the information that is needed for the different 'tolerance-situations' from the validation-experiments that will be organized during the coming years. Effects of the following on accuracy and precision should be taken into account:

- ***difference between whole and ground,***
- ***influence of oven temperature (high vs low)***
- ***approaches of other ISTA Committees like germination and try to be consistent with these, if statistically valid and possible.***

ANNEX 2.3.2.1.: Analysis of ISTA Proficiency Test data

Annex 2.3.2.1.1. The data

ISTA Proficiency Test data of Trifolium inc (Whole) and Triticum aest (Ground):
. 110 labs for Whole, 102 labs for Ground
. 3 lots
. 2 reps/lot

Annex 2.3.2.1.2. Mixed effects model for the analysis

A statistical model used to describe the data is a mixed-effects model:

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + e_{ijk}$$

where:
. y_{ijk} is the moisture observed for lot i in lab j and rep k
. i = 1, 2, 3 j = 1, …, 110 or 102   k = 1,2
. μ is the intercept.
. α_i is the fixed effect of lot i.
. b_j is the random effect of lab j. The b_j are iid N(0, σ^2_{Lab}).
. $(\alpha b)_{ij}$ is the random interaction effect between lot i and lab j. The $(\alpha b)_{ij}$ are iid N(0, $\sigma^2_{Lab \times Lot}$).
. e_{ijk} are the residuals. They are iid N(0, σ^2).

Annex 2.3.2.1.3. Repeatability, Reproducibility

We can determine the repeatability S_r^2 and the reproducibility S_R^2 from the estimates given after fitting the mixed-effects model above:

. Repeatability reflects the uncertainty among replications of measurements made on a given lot by a given laboratory: $S_r^2 = \sigma^2$

. Reproducibility reflects the uncertainty among laboratories measuring the same lot:

$S_R^2 = \sigma_{Lab}^2 + \sigma_{Lab\,\mathrm{x}\,Lot}^2$

Annex 2.3.2.1.4. Results

Whole:

Variance component estimates:

σ_{Lab}^2	0.2481
$\sigma_{Lab\,\mathrm{x}\,Lot}^2$	0.2357
σ^2	0.0459

Repeatability: S^2_r = 0.0459
Reproducibility: S^2_R = 0.4838

Ground:

Variance component estimates:

σ_{Lab}^2	0.3253
$\sigma_{Lab\,\mathrm{x}\,Lot}^2$	0.2126
σ^2	0.1487

Repeatability: S^2_r = 0.1487
Reproducibility: S^2_R = 0.5379

Reproducibility is similar whereas repeatability is smaller for Whole.

ANNEX 2.3.2.2.: Analysis of datasets from Jette Nydam

Annex 2.3.2.2.1. The data

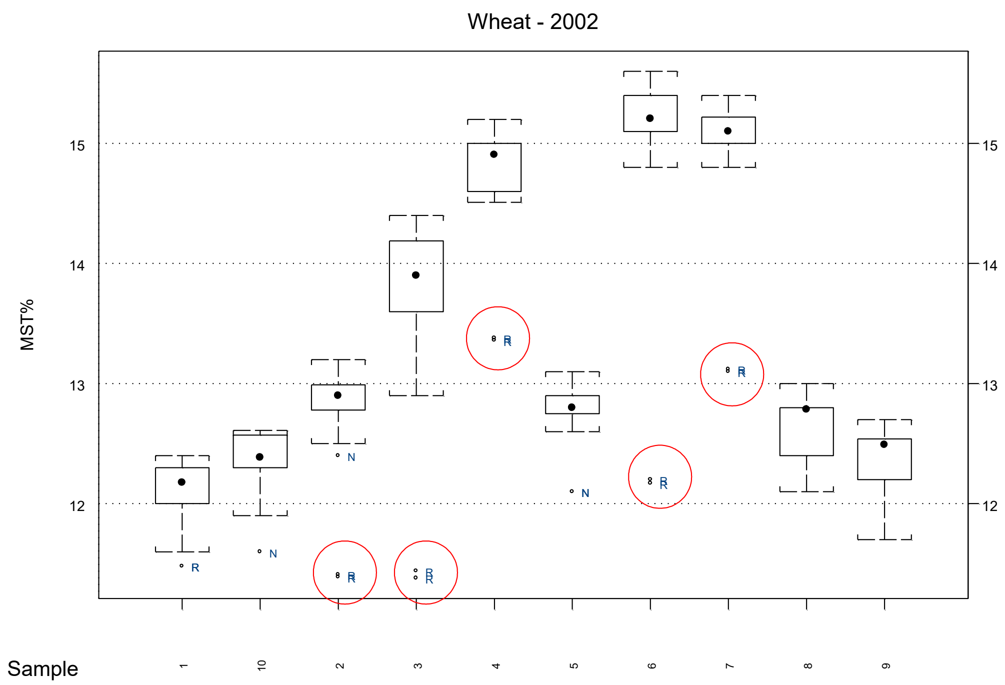
Ring test data performed within the European Grain network:

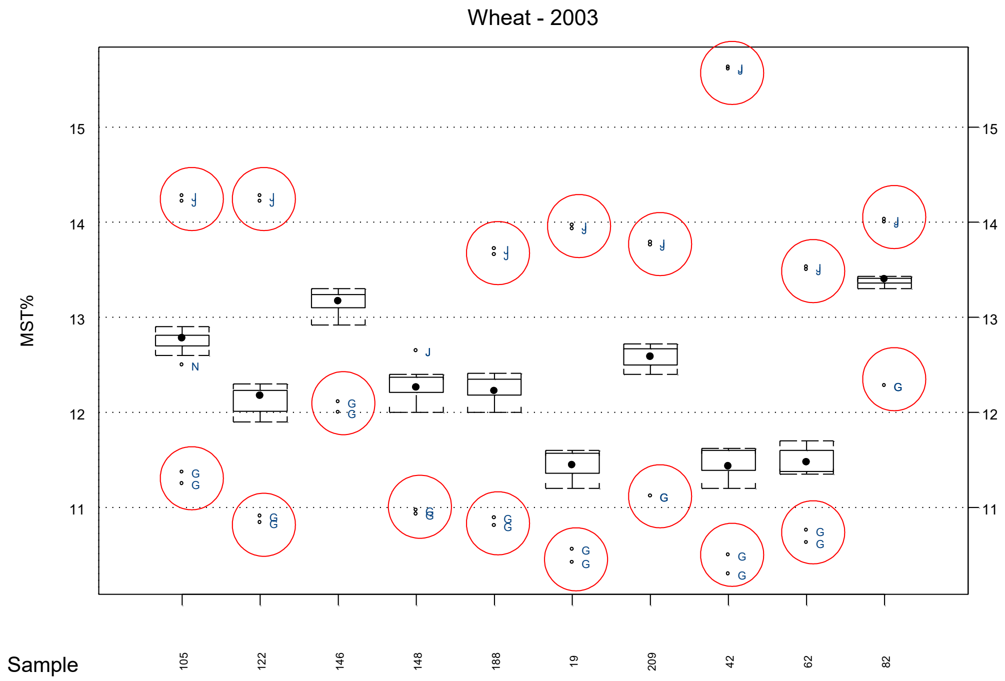
. 4 years of data for wheat and one year for Barley (2002).

. 16 labs participated in the test for wheat (some of these participated in more than 2 years), 7 labs participated in the test for barley (Labs L and S were removed from the analysis as recommended by J.Nydam) .

. Each year, 10 samples were analyzed with 2 replications.

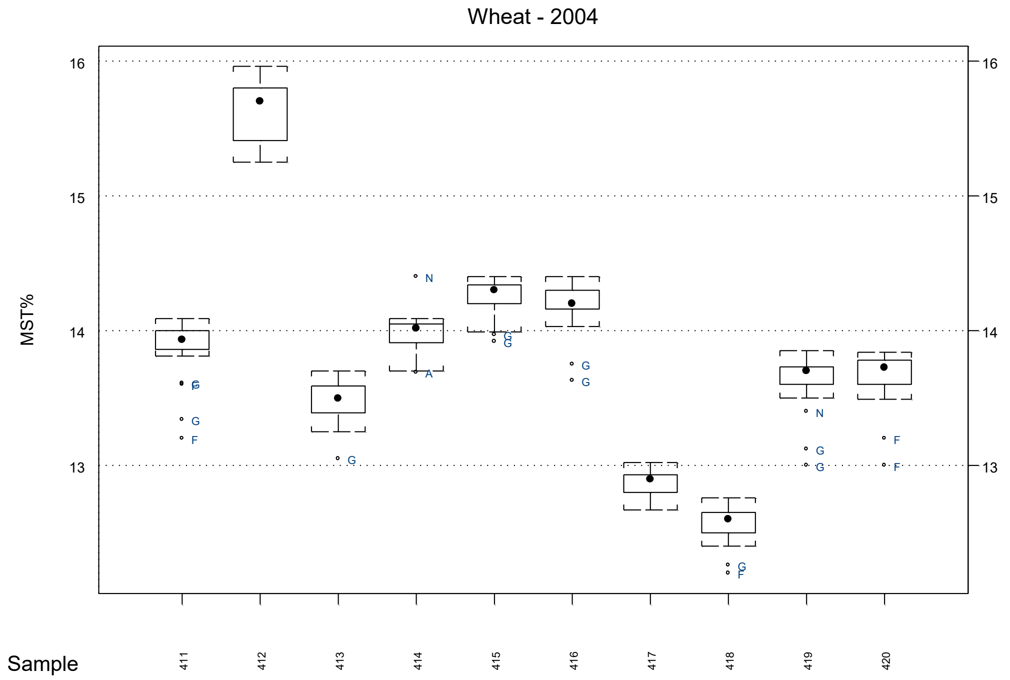
The following side-by-side boxplots exhibit the distribution of the results from all the labs each year and for each sample:

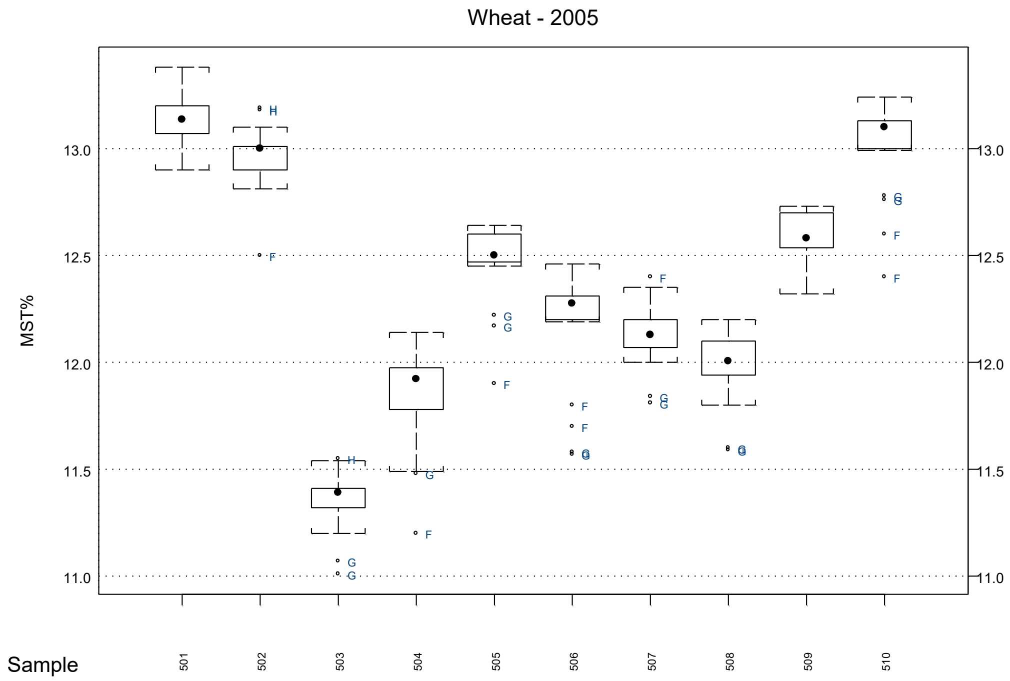

Wheat - 2002
MST%
Sample
1
10
2
3
4
5
6
7
8
9
12
13
14
15

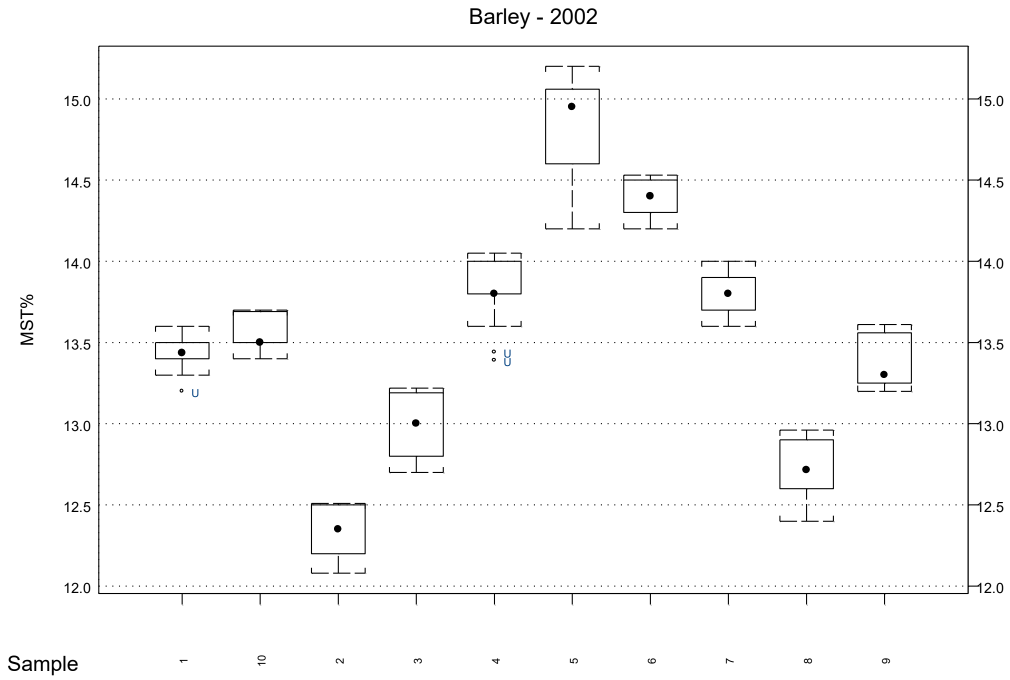

Wheat - 2003
MST%
Sample
105
122
146
148
188
19
209
42
62
82
11
12
13
14
15

INTERNATIONAL SEED TESTING ASSOCIATION (ISTA)

Secretariat, Zürichstrasse 50, 8303 Bassersdorf, CH-Switzerland
Phone: +41-44-838 60 00, Fax: +41-44-838 60 01, Email: ista.office@ista.ch, http://www.seedtest.org







From these boxplots, we can see that some labs exhibit for wheat large outside values for some years, which might affect the repeatability and reproducibility estimation.

Thus, it has been decided to remove the following Lab x Year combinations from the analysis of wheat data:
. Lab R, 2002
. Labs J and G, 2003

Annex 2.3.2.2.2. Mixed effects model for the analysis
A statistical model used to describe the data is a mixed-effects model:

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + e_{ijk}$$

where:
. y_{ijk} is the moisture observed for sample i in the lab x year combination j and rep k
. μ is the intercept.
. α_i is the fixed effect of sample i.
. b_j is the random effect of the lab x year combination j. The b_j are iid N(0, $\sigma^2_{Lab\_Yr}$).
. $(\alpha b)_{ij}$ is the random interaction effect between sample i and lab x year combination j. The $(\alpha b)_{ij}$ are iid N(0, $\sigma^2_{Samp \times Lab\_Yr}$).
. e_{ijk} are the residuals. They are iid N(0, σ^2).

Annex 2.3.2.2.3. Repeatability, Reproducibility
We can determine the repeatability S_r^2 and the reproducibility S_R^2 from the estimates given after fitting the mixed-effects model above:
. Repeatability reflects the uncertainty among replications of measurements made on a given sample by a given laboratory in a given year: $S_r^2 = \sigma^2$
. Reproducibility reflects the uncertainty among laboratories measuring the same sample:
$S_R^2 = \sigma^2_{Lab\_Yr} + \sigma^2_{Samp \times Lab\_Yr}$

Annex 2.3.2.2.4. Results

Wheat:

Variance component estimates:

$\sigma^2_{Lab\_Yr}$	0.0207
$\sigma^2_{Samp \times Lab\_Yr}$	0.0095
σ^2	0.0056

Repeatability: $S^2_r = 0.0056$
Reproducibility: $S^2_R = 0.0302$

Barley:

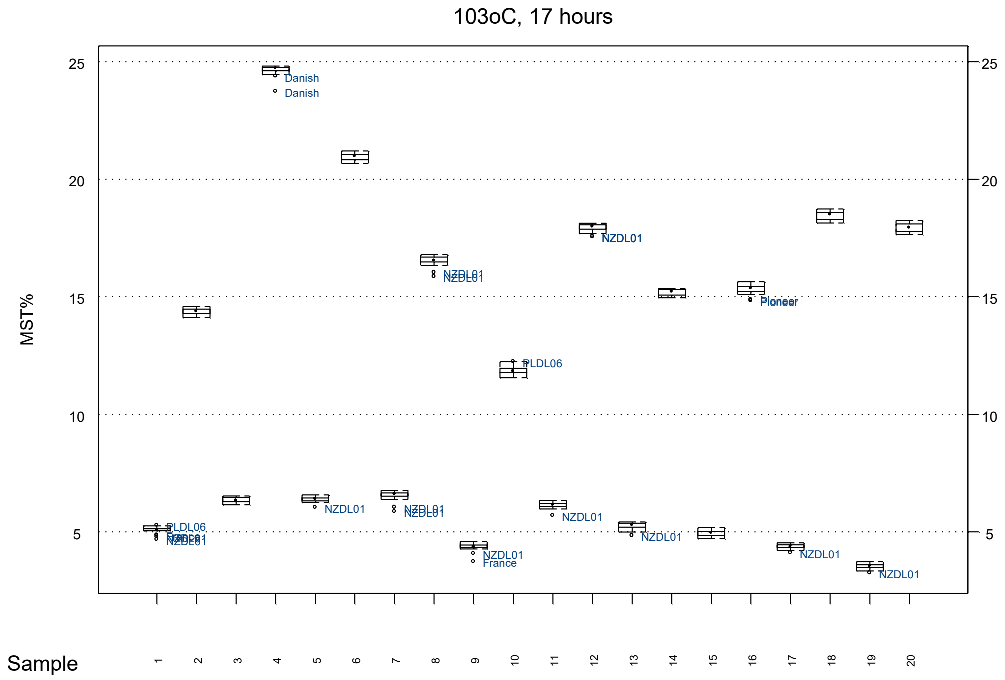
Variance component estimates:

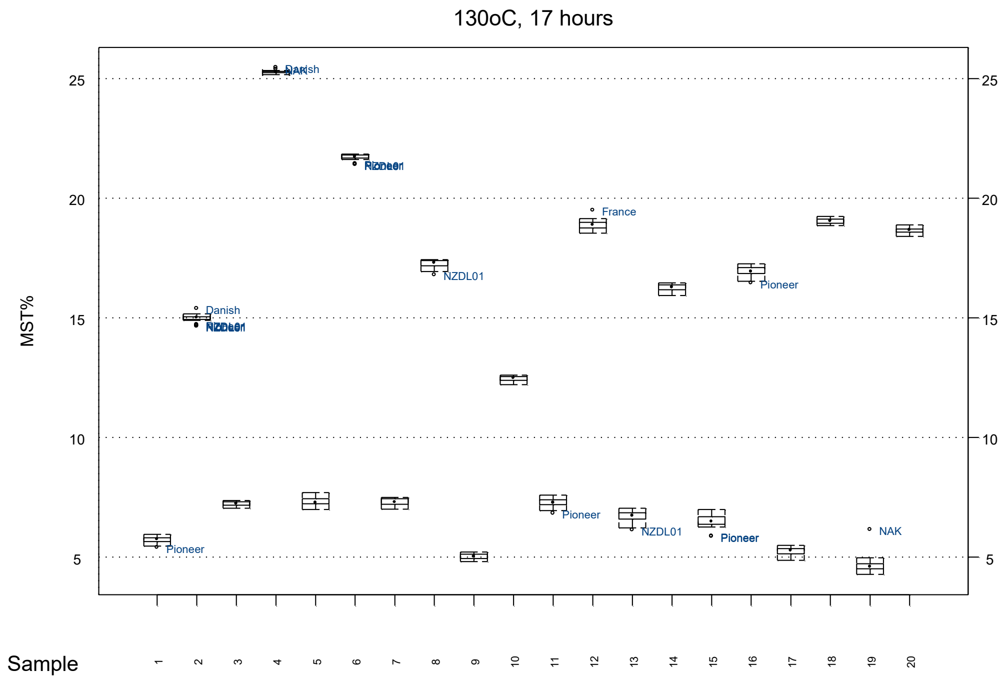
$\sigma^2_{Lab\_Yr}$	0.0156
$\sigma^2_{Samp \times Lab\_Yr}$	0.0123
σ^2	0.0057

Repeatability: $S^2_r = 0.0057$
Reproducibility: $S^2_R = 0.0280$

Repeatability and reproducibility are both similar for wheat and barley.

ANNEX 2.3.2.3: Analysis of 103°C-17 hours and 130°C-17 hours datasets from comparative testing
Annex 2.3.2.3.1. The data
10 labs analyzed 20 samples with 2 replications.
The following side-by-side boxplots exhibit the distribution of the results from all the labs and for each sample:





No large outside values are revealed by these boxplots.

Annex 2.3.2.3.2. Mixed effects model for the analysis
A statistical model used to describe the data is a mixed-effects model:

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + e_{ijk}$$

where:

. y_{ijk} is the moisture observed for sample i in lab j and rep k
. i = 1, …, 20 j = 1, …, 10 k = 1,2
. μ is the intercept.
. α_i is the fixed effect of sample i.
. b_j is the random effect of lab j. The b_j are iid N(0, σ^2_{Lab}).
. $(\alpha b)_{ij}$ is the random interaction effect between sample i and lab j. The $(\alpha b)_{ij}$ are iid N(0, $\sigma^2_{Samp\,x\,Lab}$).
. e_{ijk} are the residuals. They are iid N(0, σ^2).

Annex 2.3.2.3.3. Repeatability, Reproducibility
We can determine the repeatability S^2_r and the reproducibility S^2_R from the estimates given after fitting the mixed-effects model above:
. Repeatability reflects the uncertainty among replications of measurements made on a given lot by a given laboratory: $S^2_r = \sigma^2$
. Reproducibility reflects the uncertainty among laboratories measuring the same lot:

$$S^2_R = \sigma^2_{Lab} + \sigma^2_{Samp\,x\,Lab}$$

Annex 2.3.2.3.4. Results

103°C–17 hours

Variance component estimates:

σ^2_{Lab}	0.0129
$\sigma^2_{Samp\,x\,Lab}$	0.0106
σ^2	0.0063

Repeatability: S^2_r = 0.0063
Reproducibility: S^2_R = 0.0235

130°C-17 hours

Variance component estimates:

σ^2_{Lab}	0.0161
$\sigma^2_{Samp\,x\,Lab}$	0.0123
σ^2	0.0095

Repeatability: S^2_r = 0.0095
Reproducibility: S^2_R = 0.0284

Repeatability and reproducibility are both similar for 103°C and 130°C.