Interlaboratory Comparison Measurement EstOil-5

Final Report

09.11.2009

Report Compiled by

Lauri Jalukse

Ivo Leito

Tiina Kukk

This report is available at the website of UT at http://www.ut.ee/katsekoda/ILC/

Testing Centre, University of Tartu

Werol Tehased Ltd

Tartu 2009

Table of Contents

1 The Aim of the Intercomparison

The aim of the EstOil-5 intercomparison was to allow the participating laboratories to assess their performance in determining seven edible oil parameters: moisture content, free fatty acids content (below FFA), peroxide value (below PV), saponification value (below SAPV), beta-sitosterol content (below B-SITO), phosphorus content (below P) and erucic acid content (below EA). All parameters except P were determined in refined rapeseed oil (P was determined in crude rapeseed oil). This is the fifth intercomparison of this series. The previous intercomparisons: EstOil-1[1], EstOil-2[2], EstOil-3[3] and EstOil-4[4] took place in 2005, 2006, 2007 and 2008 respectively.

2 Organization

2.1 General

The intercomparison measurement was organized jointly by the Testing Centre of University of Tartu (below UT) and Werol Tehased Ltd (below WT). See Table 1 for the detailed contact information of the organizers.

Table 1. Contact Information of the Organizers.

University of Tartu, Testing Centre http://www.ut.ee/katsekoda Ravila 14a, 50411 Tartu, Estonia Phone: +372 51 84 176 Fax: +372 737 5264 E-mail: lauri.jalukse@ut.ee	Werol Tehased Ltd Painküla 48422, Jõgevamaa, Estonia Phone: +372 77 68 233 Fax: +372 77 68 220 E-mail: tiina.kukk@peregrupp.ee

This report was compiled jointly by UT and WT and is publicly available via the website of UT at http://www.ut.ee/katsekoda/ILC/. The participants are listed in this report but the results are presented in random order, so that the results cannot be traced back to the participants. Every participant will receive a private letter revealing his/her result number and permitting assessment of performance.

2.2 The Samples

The oil samples were prepared and distributed by WT. The samples were refined rapeseed oil (moisture content, FFA content, PV, SAPV, B-SITO content, EA content) and crude rapeseed oil (P content) samples of approximately 100 ml in gas-tight (sealed) amber glass bottles. The samples were prepared from a single bulk of oil that was well mixed before filling the bottles. The bottles were filled and closed during a short time (around 30 seconds per bottle). The laboratories got random bottles from the pool of bottles. The first and last bottles were not distributed.

[1] http://www.ut.ee/katsekoda/ILC/Estoil/Estoil_1_rep_Final.pdf

[2] http://www.ut.ee/katsekoda/ILC/Estoil/EstOil-2_2006_final_report.pdf

[3] http://www.ut.ee/katsekoda/ILC/Estoil/EstOil-3_final_report.pdf

[4] http://www.ut.ee/katsekoda/ILC/Estoil/EstOil-4_Final_Report.pdf

2.3 Data Treatment

The evaluation of participant data was done at UT according to the ISO Guide 43-1.[5] The z-score approach was used. The z-score for a particular measured value of a participant is calculated according to the following equation:

$$z = \frac{x - x_c}{s}, \tag{1}$$

where x is the participant's value, x_c is the consensus value and s is the target standard deviation. The consensus values are found as the mean values after elimination of outliers. The target standard deviation in our case is found as the real standard deviation of the participant values after elimination of outliers.

Elimination of outliers was done using the Grubbs test.[6] When applying the Grubbs test to a dataset with relatively high spread of values then extremely low values will often be retained by the test, because the high spread of the values leads to limits that will allow even negative values to be retained. At the same time it is clearly unreasonable not to eliminate values that are many times lower than the rest. This problem was tackled by applying the Grubbs test in two steps:

(1) First the full iterative Grubbs procedure was carried out on the results as presented. Any outliers were eliminated.

(2) Then logarithms were calculated from those results that were not eliminated during the first step and the logarithms of the results were subjected to a second iterative Grubbs procedure.

The moisture content was specified in the invitations to EstOil-5 as to be determined by the Karl Fischer procedure. Therefore we did not to include the results of gravimetric (heating loss) moisture determination into the data set that was used for calculation of the consensus value and target standard deviation. Nevertheless, the results were retained for assessment of participant performance.

Assessment of participant performance was carried out in two ways.

(1) Absolute values of z-scores (|z| values) are used for assessing the acceptability of the results as described in Table 2.

Table 2. Assessment of Acceptability of the Results Using z-Scores.

|z| Value	Acceptability of the Result	Required Action
$\|z\| \leq 2$	Acceptable result	No action is required
$2 < \|z\| < 3$	Doubtful result	Preventive action is required
$\|z\| \geq 3$	Unacceptable result	Corrective action is required

(2) Pairwise E_n values between participants presented as tables.

This is done using the E_n numbers as described in ISO Guide 43-1:[5]

$$E_n = \frac{C_{lab1} - C_{lab2}}{\sqrt{U_{lab1}^2 + U_{lab2}^2}} \ . \tag{2}$$

[5] ISO Guide 43-1 *Proficiency Testing by Interlaboratory Comparisons. Part 1: Development and Operation of Proficiency Testing Schemes*, ISO/IEC 1997.

[6] *AOAC Official Methods of Analysis*, Appendix D; AOAC, 1995.

where C_{lab1} and C_{lab2} are the results of the two laboratories that are compared and U_{lab1} and U_{lab2} are their expanded uncertainties. Equation 2 is adequate, if between-sample variability is significantly (more than 5 times) lower than between-participant variability. If not, the between-sample variability has to be taken into account and the E_n value is found as follows:

$$E_n = \frac{C_{lab1} - C_{lab2}}{\sqrt{U_{lab1}^2 + U_{lab2}^2 + (t_{95}(df) \cdot s_s)^2}} \ . \quad (3)$$

where s_s – is the between-sample standard deviation and $t_{95}(df)$ is the student coefficient at 95% confidence level with df degrees of freedom. Agreement between two results is considered acceptable if $|E_n| \leq 1$. Participants who did not report uncertainties for their results were excluded from the pair-wise comparisons.

In addition to the above-described data treatment schemes of the ISO Guide 43-1 we similarly to EstOil-3 and EstOil-4 carried out data treatment according to the "robust statistics" approach,[7] which is presented in Annex 1. This approach permits to avoid some of the problems of the two standard approaches presented above. Since this approach was not announced in the invitation to the intercomparison, the Annex 1 remains informatory only.

[7] Jörg W. Müller, *J. Res. Natl. Inst. Stand. Technol.* **2000**, 105, 551-555

3 Participants

Invitations were sent to a number of European laboratories. The participants are listed in Table 3.

Table 3. Participants to EstOil-5.

Institution	Country
Herkon	Bosnia
Euro Inspekt d.o.o.	Bosnia
Euroinspekt Croatiakontrola d.o.o.	Croatia
Public Health Institute	Croatia
Panchris Animal Premix ltd	Cyprus
Laboratory of Institute of Chemistry, University of Tartu	Estonia
Laboratory of Werol Tehased Ltd	Estonia
Testing Centre of University of Tartu	Estonia
Agricultural Research Centre	Estonia
AS Biodiesel Paldiski laboratory	Estonia
Chemical Analytical Laboratories "M Galanakis"	Greece
Union of Agricultural Cooperatives of Iraklio	Greece
Central Agricultural Office, Food Analytical National Reference Laboratory of Food and Feed Safety Directorate	Hungary
AS "Latvenergo" Chemical laboratory	Latvia
Laboratoire Officiel d'Analyses et de Recherches Chimiques	Morocco
Powiatowa Stacja Sanitarno-Epidemiologiczna	Poland
UOKiK Laboratorium Kontrolno-Analityczne w Olsztynie	Poland
Regionálny úrad verejného zdravotníctva so sídlom v Prešove	Slovakia
EL SPOL. S R.O.	Slovakia
Zavod Za Zdravstveno Varstvo Maribor-Inštitut Za Varstvo Okolja	Slovenia
Petrol d.d. , Ljubljana - Laboratory Petrol	Slovenia
SGS Española de Control - Laboratorio Agridiv Avda. Santa Clara de Cuba – Pol. Ind. Sta. Clara	Spain

4 Results

4.1 Results of the Participants

Results of the participants are presented in Table 4. The results are presented with the same number of decimal digits as given by the participants. Participants who presented their results in units other than those requested were asked to make the unit conversion themselves.

Table 4. Participant Results together with the Expanded Uncertainties and the Derived Consensus Values.

	Moisture Content			Free Fatty Acid Content		
Lab number[a]	Result[b] ppm[d]	Uncertainty ppm	k[c]	Result[b] %	Uncertainty %	k[c]
1						
2	392	38	2	0.033	0.006	2
3						
4	390.7	16.1	95.45%			
5	388.00	33.76	2			
6	*500.66*	*3.01*	*2*	**0.11**	0.00022	2
7	**350**	20	2	0.02		
8				0.029	0.001	2
9						
10				0.05		
11				0.029	0.002	2
12	**449.9**	0.064	2	0.069	0.03	2
13				0.036	0.04	2
14						
15	397.2	40	2			
16						
17				0.025	0.001	1.96
18	400	40.0	2	0.04	0.004	2
19						
20	382.61			0.043		
21	*165*			0.03		
22						
Consensus value	**391.8**			**0.037**		

Lab number[a]	Peroxide value Result[b] meqO2/kg	Uncertainty meqO2/kg	k[c]	Phosphorus Content in crude rapeseed oil Result[b] ppm[d]	Uncertainty ppm	k[c]
1						
2	2.5	1.4	2	127	16	2
3						
4						
5						
6	**10.9**	0.31	2			
7	3.6			145		
8	1.66	0.17	2			
9						
10	2.93					
11	2.67	0.08	2			
12	1.34	0.1	2	**399.84**		
13						
14	1.70	0.27	2			
15						
16						
17	2.43	0.16	1.96			
18	3.34	0.17	2			
19	2.37			143.62		
20	1.86			129.76		
21	0.83			136.5		
22				140.2	21	2
Consensus value	**2.27**			**137.0**		

Lab number[a]	Saponification value Result[b] mg/g	Uncertainty mg/g	k[c]	Beta-Sitosterol content Result[b] ppm[d]	Uncertainty ppm	k[c]
1				3867	702	95.45%
2	187.9	4.7	2			
3	192	2	2	**3080**	616	2
4						
5						
6						
7						
8	189	2.1	2			
9						
10						
11						
12	192.98	0.8	2			
13	184.7	0.9	2			
14						
15						
16						
17				3923	749	1.96
18				3904	78	2
19						
20	182.44					
21	186.2			3988		
22						
Consensus value	**187.9**			**3920.5**		

Lab number[a]	Erucic acid content Result[b] %	Uncertainty %	k[c]
1			
2			
3			
4			
5			
6			
7			
8	0.26	0.2	2
9	0.30		
10			
11			
12			
13			
14			
15	0.30	0.05	2
16	0.284	0.012	2
17			
18	0.26	0.013	2
19	0.27		
20			
21	0.3		
22			
Consensus value	**0.282**		

[a] The participating laboratories are given numbers in random order that is different from the order given in Table 3. [b] Outlying results according to the ordinary Grubbs test and Grubbs test with log values are marked in red and blue, respectively. [c] Coverage factor or confidence level, as provided by the participants. [d] Although ppm is not an SI unit, it was decided to present the results in this unit respecting the established practice. ppm is identical to mg/kg. [e] Moisture content values that are found by the heating loss method are given in *italic*.

Two laboratories were found outlying in moisture content measurement (only laboratories using the Karl Fischer method were considered), one laboratory in FFA measurement, one in P content measurement, none in SAPV measurement, one in B-SITO measurement, and none in EA content measurement (see Table 4) according to the Grubbs tests.

Peroxide value (PV) is a special parameter because of its instability. As Figure 4 demonstrates, there is no trend in the PV values obtained by different laboratories at different times. According to the experiments carried out at our laboratory the PV increases in a closed bottle kept at room temperature by ca 0.06 meq/kg per ten days. Given that in the worst cases the measurements were done within a 16-day time window the reasonable increase of PV over time would be in the range of 0.1-0.2 meq/kg. It is possible that with some bottles the increase was faster, but in any case, the rate of PV increase is too vaguely defined to allow any meaningful correction. Similarly, we did not consider it justified to use the Grubbs test – a purely statistical method – for outlier detection in the case of PV. So the data of all laboratories with the exception of laboratory No 6 were used for the consensus value and target standard deviation calculation. Laboratory No 6 reported result, which is around five times larger than the consensus value and almost three times larger than the next largest value. This value cannot be explained by any temporal change in the oil sample.

Full information about consensus values and target standard deviations is presented in Table 5 (the respective data of the EstOil-1, EstOil-2, EstOil-3 and EstOil-4 intercomparisons are also given for reference).

Table 5. Consensus Values and Target Standard Deviations of Interlaboratory Comparison Measurements EstOil-5, EstOil-4, EstOil-3, EstOil-2 and EstOil-1.

Moisture content	FFA content	PV	P content	SAPV	B-SITO content	EA content	
ppm	%	meq/kg	ppm	mg/g	ppm	%	Unit
			EstOil-5				
391.8	0.037	2.27	137.0	187.9	3920.5	0.282	consensus value
391.4	0.033	2.40	138.4	187.9	3913.5	0.284	consensus value (median)
6.3	0.014	0.82	7.4	3.8	50.7	0.019	target standard deviation
5.4	0.004	0.38	5.0	2.4	43.2	0.012	target standard deviation (median)
0.0	0.0005	0.05	1.0	1.3	101.2	0.003	between-sample variability[a]
2%	37%	36%	5%	2.0%	1%	7%	relative target standard deviation
1%	13%	16%	4%	1.3%	1%	4%	relative target standard deviation (median)
0.0%	1.3%	2.2%	0.8%	0.7%	2.6%	1.2%	relative between-sample variability
0.1%	10.1%	5.8%	1.0%	0.0%	0.2%	0.7%	relative difference between two consensus values
			EstOil-4				
373.7	0.036	1.71	130.5	190.2	3532.3	0.097	consensus value
365.0	0.033	1.51	123.0	191.0	3688.6	0.099	consensus value (median)
20.6	0.017	0.79	18.9	2.9	427.5	0.007	target standard deviation
13.0	0.004	0.18	8.4	0.7	132.9	0.005	target standard deviation (median)
0.5	0.002	0.00	2.2	1.0	27.0	0.006	between-sample variability[a]
6%	48%	46%	15%	1.5%	12%	7%	relative target standard deviation
4%	12%	12%	7%	0.4%	4%	5%	relative target standard deviation (median)
0.1%	4.2%	0.0%	1.7%	0.5%	0.8%	6.2%	relative between-sample variability
2.3%	8.5%	12.1%	5.7%	0.4%	4.4%	1.7%	relative difference between two consensus values
			EstOil-3				
81.6	0.031	14.07	125.3	189.7	3877.4		consensus value
101.2	0.029	14.00	138.1	190.1	3820.5		consensus value (median)
22.4	0.013	3.60	55.4	2.3	423.5		target standard deviation
16.7	0.003	0.45	13.6	1.0	143.7		target standard deviation (median)
2.9	0.0000	0.20	2.7	1.0	59.1		between-sample variability[a]
27%	42%	26%	44%	1.2%	11%		relative target standard deviation
16%	11%	3%	10%	0.5%	4%		relative target standard deviation (median)
3.6%	0.0%	1.4%	2.2%	0.5%	1.5%		relative between-sample variability
23.9%	7.7%	0.5%	10.2%	0.2%	1.5%		relative difference between two consensus values
			EstOil-2				
87.7	0.03	6.9	124.2				consensus value
19.1	0.01	1.9	31.9				target standard deviation
6.4	0.0003	0.03	1.9				between-sample variability[a]
22%	37%	27%	26%				relative target standard deviation
7.3%	1.1%	0.4%	1.5%				relative between-sample variability
			EstOil-1				
362.5	0.07						consensus value
32.9	0.01						target standard deviation
6.0	0.0004						between-sample variability[a]
9%	19%						relative target standard deviation
1.7%	0.5%						relative between-sample variability

[a] Given at standard deviation level. See section 4.2 for more information.

The z-scores are calculated according to equation 1 and are presented in Table 6.

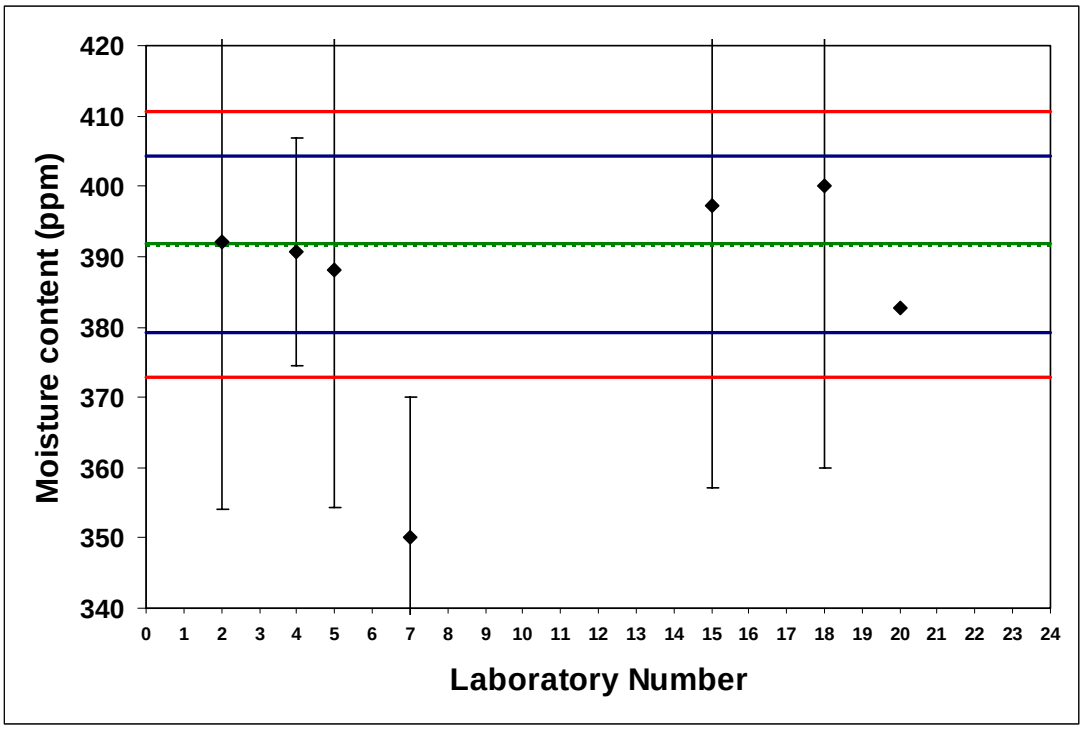
Table 6. Participant z-Scores.

Lab number[a]	Moisture content	FFA	PV	P content	SAPV	B-SITO content	ERUC content
1						-1.1	
2	0.0	-0.3	0.3	-1.4	0.0		
3					1.1	-16.6	
4	-0.2						
5	-0.6						
6	17.4	5.4	10.5				
7	-6.7	-1.2	1.6	1.1			
8		-0.6	-0.7		0.3		-1.2
9							1.0
10		1.0	0.8				
11		-0.6	0.5				
12	9.3	2.4	-1.1	35.7	1.3		
13		-0.1			-0.8		
14			-0.7				
15	0.9						1.0
16							0.1
17		-0.9	0.2			0.0	
18	1.3	0.2	1.3			-0.3	-1.2
19			0.1	0.9			-0.6
20	-1.5	0.5	-0.5	-1.0	-1.4		
21	-36.2	-0.5	-1.8	-0.1	-0.4	1.3	1.0
22				0.4			

[a] The participating laboratories are given in random order that is different from the order given in Table 3 but is identical to the order given in Table 4. [b] According to the ISO Guide 43-1: acceptable result is marked in green, doubtful result in yellow and unacceptable result in red.

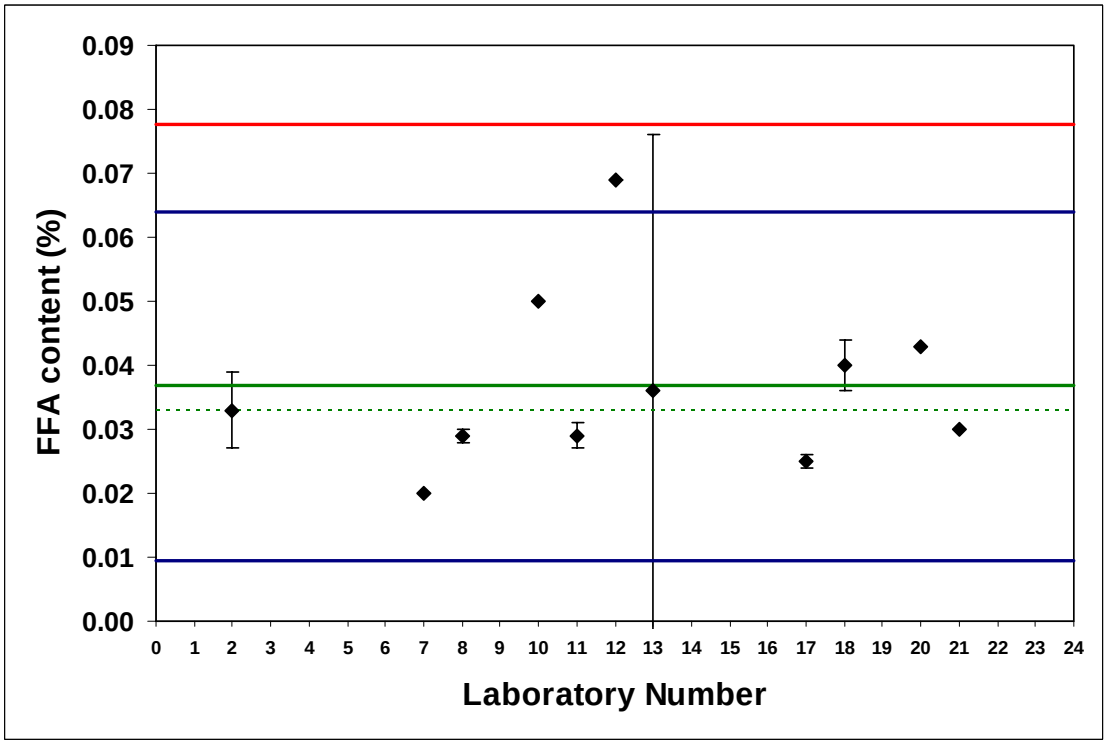
The results are presented in graphical form in the Figures below:

Figure 1. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] Moisture Content Measurement[b].



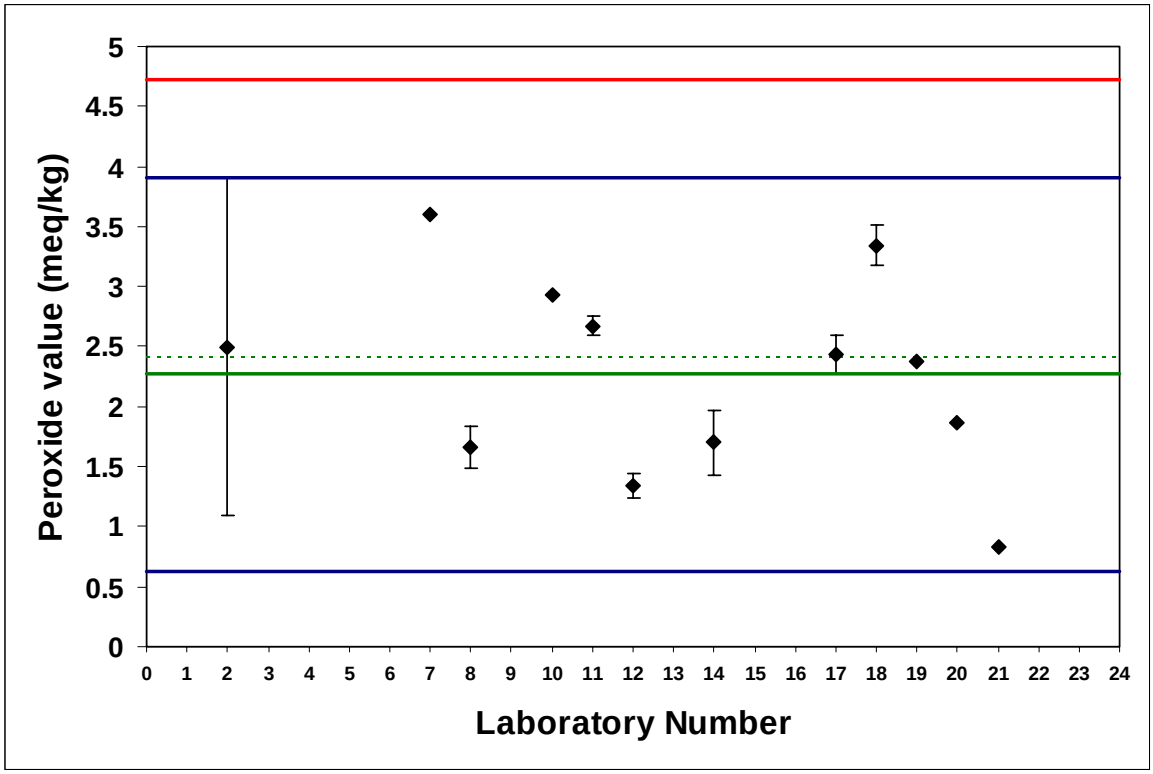
[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b] Results of the laboratories 6, 12 and 21 are out of the figure range (see the Table 4).

Figure 2. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] FFA Content Measurement[b].



[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b] Result of the laboratory 6 is out of the figure range (see the Table 4).

Figure 3. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] PV Measurement[b].



[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b] Result of the laboratory 6 is out of the figure range (see the Table 4).

Figure 4. Dependence of the Results of PV Measurement on Measurement Date.[a]

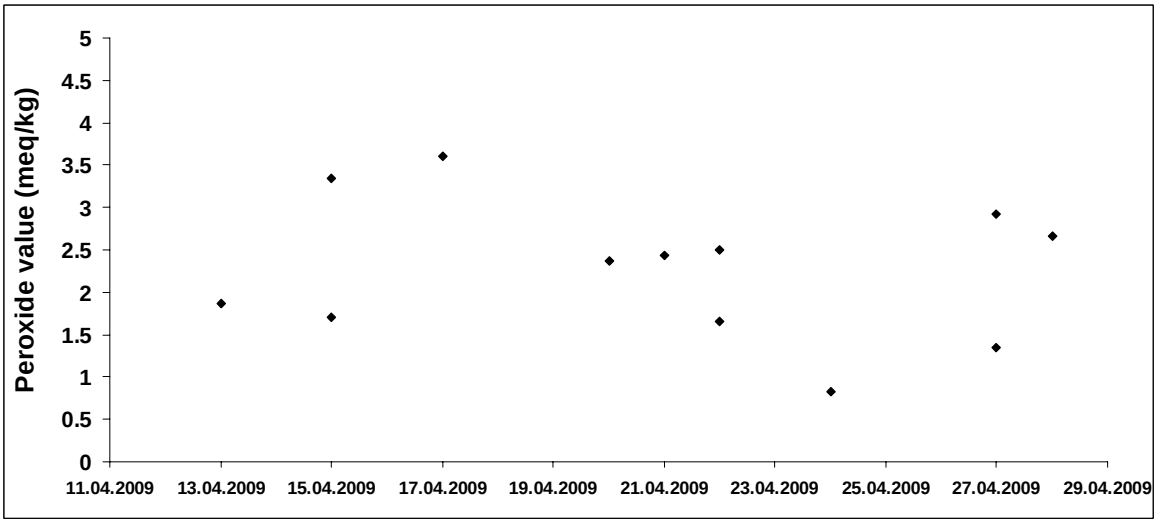
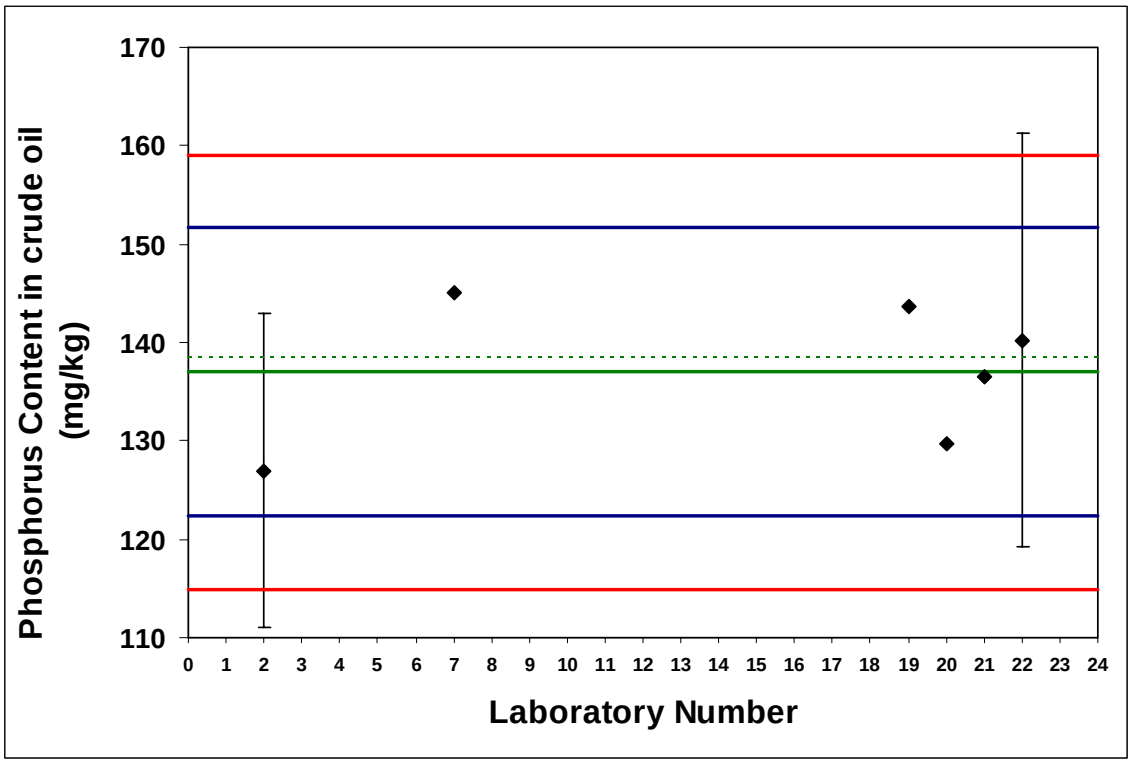
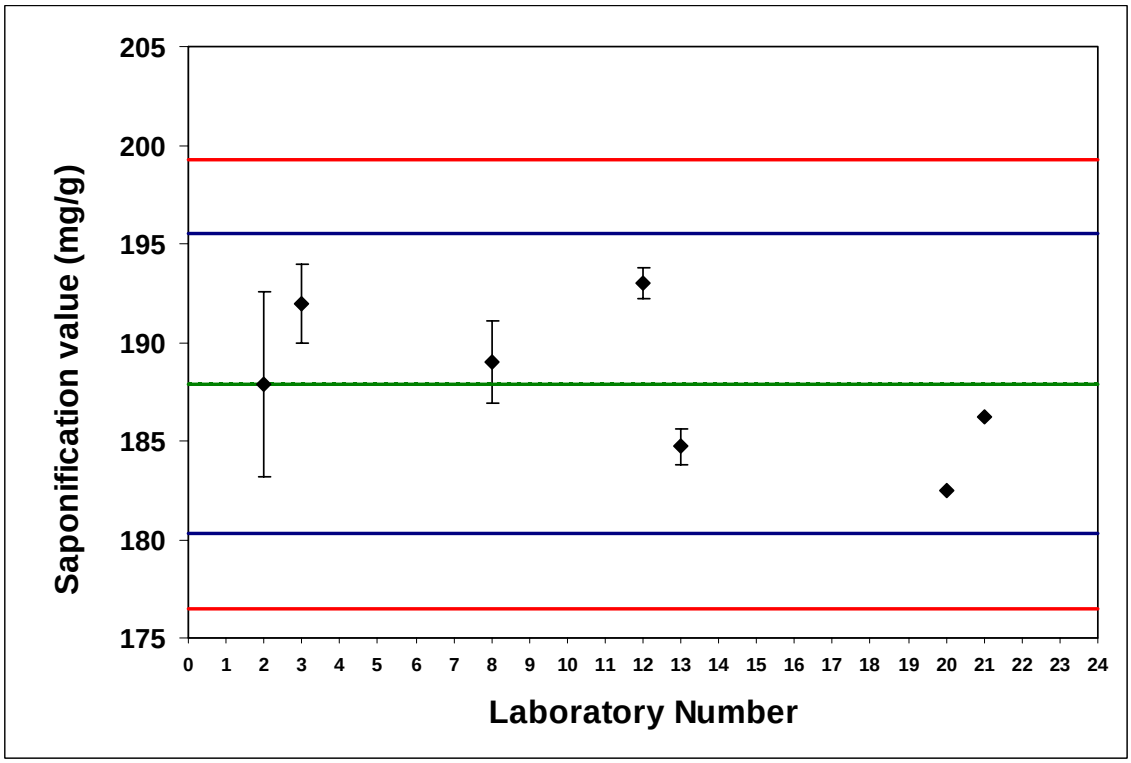


Figure 5. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] P Content Measurement[b].



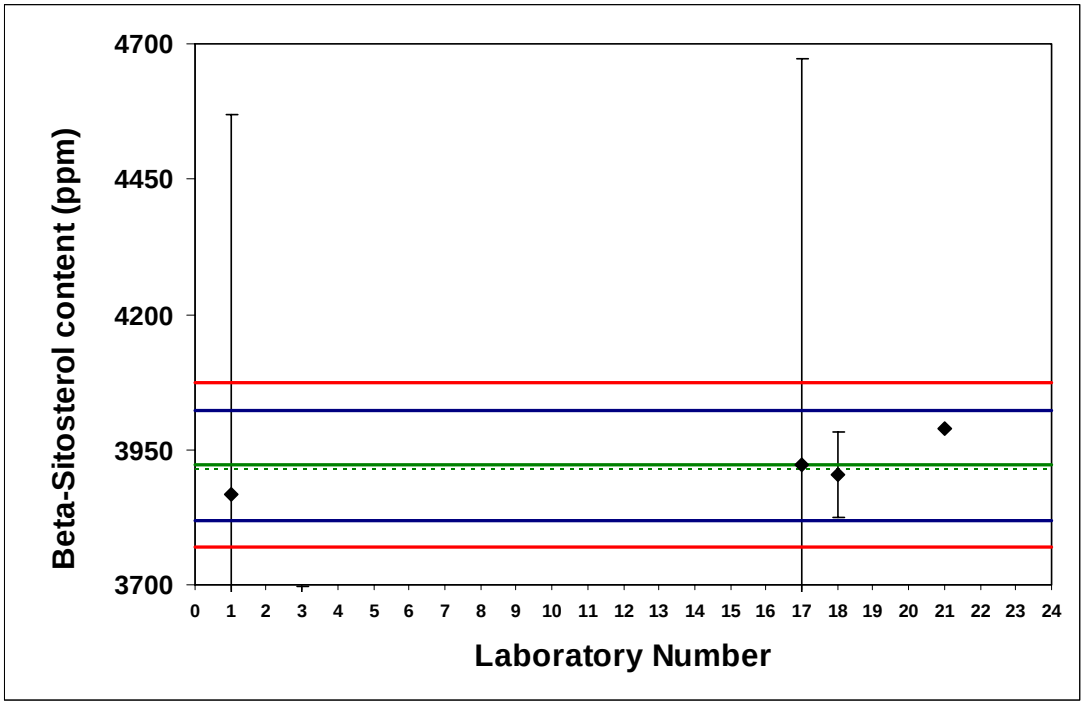
[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b] Result of the laboratory 12 is out of the figure range (see the Table 4).

Figure 6. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] SAPV Measurement.



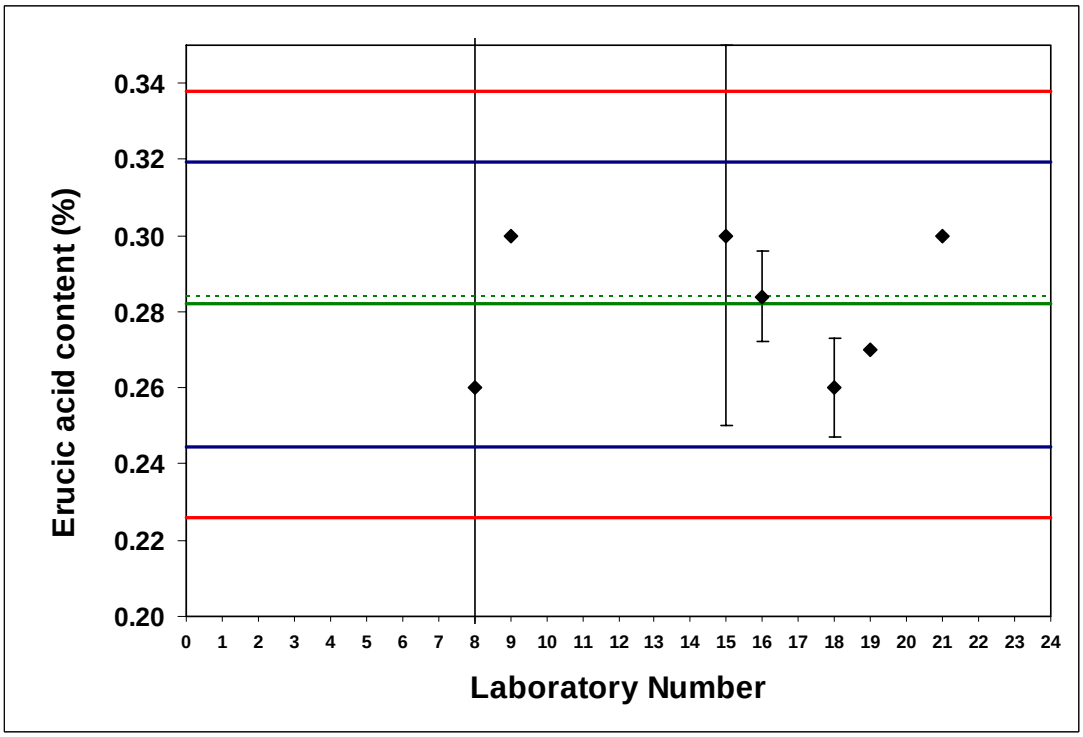
[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines.

Figure 7. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] B-SITO Measurement[b].



[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b] Result of the laboratory 3 is out of the figure range (see the Table 4).

Figure 8. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] EA Content Measurement.



[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines.

From Table 6 and the Figures it can be concluded that based on the z-score approach from the 61 submitted results 9 are unacceptable and 1 is doubtful.

4.2 Between-Sample Variability

Between-sample variability was determined by UT (moisture content), by WT (FFA content, PV, P content, SAPV), by Institute of Chemistry, University of Tartu (B-SITO content) and by Estonian Agricultural Research Centre (EA content) under repeatability conditions (see Table 5). The data were treated using the ANOVA technique to separate the effects of between- and within-sample variability.[8]

For moisture content the between-sample standard deviation could not be calculated because the ANOVA results indicated that the overall variability is wholly due to within-sample variability (five samples, coulometric Karl Fischer titration). This indicates that between-sample variability is negligible. For FFA content the between-sample standard deviation was 0.0005% (five samples). This is 29 times lower than the between-participant standard deviation. For PV content the between-sample standard deviation was 0.05 meq/kg (four samples). This is 16 times lower than the between-participant standard deviation. For phosphorus content the between-sample standard deviation was found 1.0 ppm (five samples), which is 7 times lower than the between-participant standard deviation. For SAPV the between-sample standard deviation was found 1.3 mg/g (five bottles), which is only 3 times lower than the between-participant standard deviation. For B-SITO the between-sample standard deviation was found 101.2 ppm, which is two times lower than the between-participant standard deviation. For the EA the between-sample standard deviation was found 0.003%, which is 6 times lower than the between-participant standard deviation.

We can conclude that in moisture content, FFA content, PV, P content and EA content measurement the between-sample variability has negligible effect on the between-participant variability. In the case of SAPV and B-SITO measurements the differences are 3 times lower and 2 times over, respectively. This between-sample variability is not negligible. Therefore with these parameters the between-sample variability is taken into account in the E_n score calculations (see section 2.3). It is important to note that this is a relative effect and is caused by the very low spread of the participant results in SAPV and especially B-SITO content measurement (see section 5.2 for discussion) and not by sample instability.

[8] The treatment was carried out as described in A.M.H., van der Veen, J. Pauwels, *Accred. Qual. Assur*., **2000**, *5*, 464-469.

5 Discussion

5.1 Elimination of Outliers Using the Two-Step Grubbs Test

Altogether 6 participant results were eliminated from consensus value and target standard deviation calculation based on the Grubbs tests. No laboratories were rejected using the second Grubbs test with logarithmic values (see section 4.1).

5.2 Assessment of Participant Results by the z-Score Approach

All in all 61 results were submitted. According to the z-score approach 51 of them (83.6%) were acceptable, 1 (1.6%) was doubtful and 9 (14.8%) were unacceptable (See Table 6).

With several measurands the large number of acceptable z-scores is caused by using the actual standard deviations as target standard deviations in z-score calculation. Even after outlier elimination the standard deviations of FFA and PV are still large being around 36 .. 37% of the consensus value. The situation with moisture, P, SAPV, B-SITO and EA content (relative target standard deviation being 1 .. 7%) is good.

The spread of the participant results in moisture, P, SAPV, B-SITO and EA content determination is good

This very low spread of the participant values also explains why the between-sample variability in these determinations is not significantly lower than the target standard deviation. Based on the between-sample variability determinations it can be stated that in all other cases the spread of the participant results is caused by large between-lab scatter of the results. This could be due to the following factors:

1. The FFA content in the samples were low. This is probably one of the main reasons for the large spread of these measurement results. Several factors that at higher analyte levels are of low significance can seriously influence results at low analyte levels. In the case of e.g. FFA content these can be

    a. too small size of sample taken for titration and correspondingly low titrant volume, leading to the high sensitivity of the result towards end-point determination;

    b. in the case of a larger sample, too low solvent volume, so that part of the sample does not dissolve;

    The content of FFA was similar to the EstOil-3 and EstOil-4 intercomparisons and the relative target standard deviations are also very similar, indicating consistent performance of the participants.

2. PV is an unstable analyte and the PV was also low.

The high spread FFA and PV results is caused first of all by the low values of the parameters in the samples.

3. Different methods were used by different participants. This can have an effect on the agreement of the participant results and concerns first of all the moisture content measurement.

Although according to the z-score approach most of the participants performed satisfactorily, it is of interest to compare the self-declared uncertainties of the participant results to the agreement between the participant results. The picture is very non-uniform. In the case of SAPV, EA and B-SITO content the situation is good. At the same time

> **in the case of moisture, FFA and PV the majority of participants have severely underestimated their uncertainties.**

For example, in FFA content measurements the uncertainty intervals of laboratories 8, 17 and 18 do not overlap in any of the two-lab pairs! This means that at least two of the three laboratories have underestimated the uncertainties of their results (and this is not the only similar set of laboratories). At the same time all these laboratories have satisfactory z-score values. The situation with PV and SAPV is similar.

Obviously, especially when having data with high spread like in the present intercomparison, the z-score approach has serious deficiencies when assessing participant performance:

(1) Uncertainties of participant results are not taken into account.

(2) The consensus values derived from the participant data are too unreliable to be used as reference values.

(3) The standard deviations of the participant data are too large and result in excessively wide acceptable result zones. Because of the unreliability of the consensus value it is also not reasonable to use a narrower predefined target standard deviation because then it is possible that laboratories obtaining correct results will have unacceptable z-scores.

All these problems would be solved by independently determined reference values for the samples. However, this would make the intercomparison too expensive. Under these circumstances a useful alternative is the pair-wise comparison of laboratory results using the E_n scores.

5.3 Pair-wise Comparison of Participant Results

The paired comparisons are presented in Tables 7 to 12.

Table 7. Comparison of the Results of Participants in Pairs: Moisture Content Measurement.

	$\lvert E_n \rvert$ value							
Lab No	**2**	**4**	**5**	**6**	**7**	**12**	**15**	**18**
2	0.0	0.0	0.1	2.9	1.0	1.5	0.1	0.1
4	0.0	0.0	0.1	6.7	1.6	3.7	0.2	0.2
5	0.1	0.1	0.0	3.3	1.0	1.8	0.2	0.2
6	2.9	6.7	3.3	0.0	7.4	16.9	2.6	2.5
7	1.0	1.6	1.0	7.4	0.0	5.0	1.1	1.1
12	1.5	3.7	1.8	16.9	5.0	0.0	1.3	1.2
15	0.1	0.2	0.2	2.6	1.1	1.3	0.0	0.0
18	0.1	0.2	0.2	2.5	1.1	1.2	0.0	0.0

[a] The numbers of participants are the same as in Table 4. [b]According to the ISO Guide 43-1: considered acceptable agreement between two results is considered acceptable - marked in green and unacceptable results in yellow.

Table 8. Comparison of the Results of Participants in Pairs: FFA Measurement.

	$\lvert E_n \rvert$ value							
Lab No	2	6	8	11	12	13	17	18
2	0.0	12.8	0.7	0.6	1.2	0.1	1.3	1.0
6	12.8	0.0	79.1	40.3	1.4	1.8	83.0	17.5
8	0.7	79.1	0.0	0.0	1.3	0.2	2.8	2.7
11	0.6	40.3	0.0	0.0	1.3	0.2	1.8	2.5
12	1.2	1.4	1.3	1.3	0.0	0.7	1.5	1.0
13	0.1	1.8	0.2	0.2	0.7	0.0	0.3	0.1
17	1.3	83.0	2.8	1.8	1.5	0.3	0.0	3.6
18	1.0	17.5	2.7	2.5	1.0	0.1	3.6	0.0

[a] The numbers of participants are the same as in Table 4. [b] According to the ISO Guide 43-1: considered acceptable agreement between two results is considered acceptable - marked in green and unacceptable results in yellow.

Table 9. Comparison of the Results of Participants in Pairs: PV Measurement.

	$\lvert E_n \rvert$ value							
Lab No	2	6	8	11	12	14	17	18
2	0.0	5.9	0.6	0.1	0.8	0.6	0.0	0.6
6	5.9	0.0	26.1	25.7	29.3	22.4	24.3	21.5
8	0.6	26.1	0.0	5.4	1.6	0.1	3.3	7.0
11	0.1	25.7	5.4	0.0	10.4	3.4	1.3	3.6
12	0.8	29.3	1.6	10.4	0.0	1.3	5.8	10.3
14	0.6	22.4	0.1	3.4	1.3	0.0	2.3	5.2
17	0.0	24.3	3.3	1.3	5.8	2.3	0.0	3.9
18	0.6	21.5	7.0	3.6	10.3	5.2	3.9	0.0

[a] The numbers of participants are the same as in Table 4. [b] According to the ISO Guide 43-1: considered acceptable agreement between two results is considered acceptable - marked in green and unacceptable results in yellow.

Table 10. Comparison of the Results of Participants in Pairs: P Content Measurement.

	$\lvert E_n \rvert$ value	
Lab No	2	22
2	0.0	0.5
22	0.5	0.0

[a] The numbers of participants are the same as in Table 4. [b] According to the ISO Guide 43-1: considered acceptable agreement between two results is considered acceptable - marked in green and unacceptable results in yellow.

Table 11. Comparison of the Results of Participants in Pairs: SAPV Measurement.

	$\lvert E_n \rvert$ value				
Lab No	2	3	8	12	13
2	0.0	0.7	0.2	0.8	0.5
3	0.7	0.0	0.6	0.2	1.7
8	0.2	0.6	0.0	0.9	1.0
12	0.8	0.2	0.9	0.0	2.1
13	0.5	1.7	1.0	2.1	0.0

[a] The numbers of participants are the same as in Table 4. [b] According to the ISO Guide 43-1: considered acceptable agreement between two results is considered acceptable - marked in green and unacceptable results in yellow. Because the between sample variability uncertainty of SAPV measurement is relatively high, it was necessary to use the modified equation (eq 3) for E_n value calculation.

Table 12. Comparison of the Results of Participants in Pairs: B-SITO Content Measurement.

	$\|E_n\|$ value			
Lab No	**1**	**3**	**17**	**18**
1	0.0	0.8	0.1	0.0
3	0.8	0.0	0.8	1.2
17	0.1	0.8	0.0	0.0
18	0.0	1.2	0.0	0.0

[a] The numbers of participants are the same as in Table 4. [b] According to the ISO Guide 43-1: considered acceptable agreement between two results is considered acceptable - marked in green and unacceptable results in yellow. Because the between sample variability uncertainty of B-SITO measurement is relatively high, it was necessary to use the modified equation (eq 3) for E_n value calculation.

Table 13. Comparison of the Results of Participants in Pairs: EA Content Measurement.

	$\|E_n\|$ value			
Lab No	**8**	**15**	**16**	**18**
8	0.0	0.2	0.1	0.0
15	0.2	0.0	0.3	0.8
16	0.1	0.3	0.0	1.4
18	0.0	0.8	1.4	0.0

[a] The numbers of participants are the same as in Table 4. [b] According to the ISO Guide 43-1: considered acceptable agreement between two results is considered acceptable - marked in green and unacceptable results in yellow.

From Tables 7 to 12 it can be seen that in SAPV, B-SITO and EA content determination between-lab agreements dominate.

The situation is distinctly different with the remaining three measurands. Disagreeing comparisons dominate: moisture content 57%, FFA 61% and PV 75% of the comparison pairs have disagreement. At the same time the z-scores of participants look good. This clearly indicates, that most of the participants should take a close look at their uncertainty estimates of these measurements. The abovementioned factors provide some guidelines. As a conclusion:

The pair-wise agreement of participant results in SAPV, B-SITO and EA content determination is satisfactory while in the determination of moisture content, FFA and PV it is unsatisfactory.

The uncertainties of the results of most participants in moisture content, FFA and PV determination have been underestimated.

6 Acknowledgments

The organizers would like to thank Mr Viktor Vabson (AS Metrosert, Division of Standards Services), Dr Olev Saks (Testing Centre, University of Tartu), Dr. Rouvim Kadis (D. I. Mendeleyev Institute for Metrology, Russia), Dr. Koit Herodes (University of Tartu, Department of Chemistry), MSc Erkki Mäeorg (Laboratory of Institute of Chemistry, University of Tartu), Mrs Ann Akk (Agricultural Research Centre) for advice and help.

7 Annex 1

The usual statistical algorithm of finding z scores[5] may give not the best estimates of z scores of the participants with several of the analytes determined in this intercomparison. The reasons for this are: (1) there are serious outliers (gross errors) among the data, (2) the results of the determinations carried out near the detection limit cannot be assumed to be normally distributed and (3) the results of different participants cannot be assumed to have the same uncertainty. Therefore the arithmetic mean may not be the ideal consensus value and z score may not be the ideal performance criterion.

Below we apply an alternative data analysis procedure based on the E_n scores using median and its uncertainty as the estimate of the consensus value and its uncertainty, respectively.[7] Arithmetic mean value is known to lack stability against outliers. Median has a significantly better statistical "robustness".

For a continuous variate C_{lab}, the median as a consensus value C_c is defined, using the (cumulative) distribution function $F(C_{lab})$, by the condition:

$$F(C_c) = \frac{1}{2} \ . \tag{4}$$

This means that one half of the observations are below and the other above the median. For sample of n ordered variables C_{lab1}, C_{lab2}, …, C_{labn}, the sample median, denoted as C_c=med{C_{labi}}, is given by (with integer k)

$$C_c = \begin{cases} C_{labk+1}, \ldots\ldots\ldots\ldots\ldots k = \dfrac{n-1}{2} \ for\ odd\ n \\ 1/2 (C_{labk} + C_{labk+1}), \ldots\ldots k = \dfrac{n}{2} \ for\ even\ n \end{cases} \tag{5}$$

Uncertainty of median is found as follows:

$$u(C_c) = D \cdot MAD \tag{6}$$

where D is defined as follows:

$$D = \frac{1.858}{\sqrt{n-1}} \tag{7}$$

and the where the value MAD is given by:

$$MAD = med\left\{\left|C_{labi} - C_c\right|\right\}, \ for\ \ i = 1, 2, \ldots, n. \tag{8}$$

The median-based consensus values for the measurands are given in Table 5.

Assessment of the results is done using the E_n numbers as described in ISO Guide 43-1:[5]

$$\left|E_n\right| = \frac{\left|C_{lab} - C_c\right|}{\sqrt{U_{lab}^2 + U_c^2}} \tag{9}$$

where C_{lab} are the results of a laboratory, C_c is the median as a consensus value and U_{lab} and U_c are the expanded uncertainties of the laboratory value and the median, respectively. Equation 9 is adequate, if between-sample variability is significantly (more than 5 times) lower than between-participant variability. If not, the between-sample variability has to be taken into account and the

E_n value is found as follows:

$$|E_n| = \frac{|C_{lab} - C_c|}{\sqrt{U_{lab}^2 + U_c^2 + (t_{95}(df) \cdot s_s)^2}} \quad (10)$$

where s_s – is the between-sample standard deviation and $t_{95}(df)$ is the student coefficient at 95% confidence level with df degrees of freedom.

Agreement between two results is considered acceptable if $|E_n| \leq 1$.

The results of this data treatment are presented in Table 14.

Table 14. Participant $|E_n|$ values according to new approach.

	$\|E_n\|$ scores[b] accdording to median approach						
Lab number[a]	Moisture content	FFA	PV	P content	SAPV	B-SITO content	EA content
1						0.1	
2	0.0	0.0	0.1	0.6	0.00		
3					0.02	1.2	
4	0.0						
5	0.1						
6	9.8	9.2	10.5				
7	1.8	1.5	1.6	0.7			
8		0.5	1.0		0.01		0.1
9							0.1
10		2.0	0.7				
11		0.5	0.4				
12	5.4	1.2	1.4	25.9	0.03		
13		0.1			0.02		
14			0.9				
15	0.1						0.1
16							0.0
17		0.9	0.0			0.0	
18	0.2	0.8	1.2			0.0	0.1
19			0.0	0.5			0.1
20	0.8	1.2	0.7	0.9	0.03		
21	21.0	0.4	2.1	0.2	0.01	0.3	0.1
22				0.1			

[a] The participating laboratories are given in random order that is different from the order given in Table 3 but is identical to the order given in Table 4. [b] According to the ISO Guide 43-1: acceptable result is marked in green and unacceptable result in yellow. The results of the participants who did not report uncertainties were assigned zero uncertainty.

The results presented in Table 14 are of informative nature for the current intercomparison round but more investigations will be performed and in the future this data treatment may be considered as the definitive one. If this decision will be made then this will be stated in the invitation to the intercomparison.