Interlaboratory Comparison Measurement EstOil-3

Final Report

02.11.2007

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 AS

Tartu 2007

Table of Contents

1 The Aim of the Intercomparison

The aim of the EstOil-3 intercomparison was to allow the participating laboratories to assess their performance in determining five edible oil parameters: moisture content, free fatty acids (below FFA) content, peroxide value (below PV), saponification value (below SAPV), beta-sitosterol content (below B-SITO) in refined rapeseed oil and phosphorus (below P) content in crude rapeseed oil. This is the third intercomparison of this series. The previous intercomaprisons: EstOil-1[1] and EstOil-2[2], took place in 2005 and 2006, respectively.

2 Organization

2.1 General

The intercomparison measurement was organized jointly by Testing Centre of University of Tartu (below UT) and Werol Tehased AS (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
Jakobi 2, 51014 Tartu, Estonia
Phone: +372 51 84 176
Fax: +372 737 5264
E-mail: lauri.jalukse@ut.ee

Werol Tehased AS

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) 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.

2.3 Data Treatment

The evaluation of participant data was done at UT according to the ISO Guide 43-1.[3] The z-score approach was used. The z-score for a particular measured value of a participant is calculated according to the following equation:

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

$$z = \frac{x - x_c}{s}, \quad (1)$$

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.[4] 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. 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) Logarithms were calculated from those results that were not eliminated during first step and the logarithms of the results were subjected to a second iterative Grubbs procedure.

In addition, special study was undertaken concerning applicability of the ISO 662 standard (gravimetry based on measurement of oil mass decrease on heating) to edible oils with low moisture content. It was found that the results obtained using this standard have high spread and systematically higher than those obtained using KF titration. Based on this it was decided not to include these results into the set that is used for calculation of the consensus value and target standard deviation. Nevertheless, the results were retained in the data treatment.

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:[3]

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

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:

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

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

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 additionally carried out data treatment according to the "robust statistics" approach,[5] 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.

[5] 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-3.

Institution	Country
Central Laboratory of Arme Forces	Albania
Euro Inspekt d.o.o.	Bosnia
Herkon	Bosnia
Euroinspekt Croatiakontrola d.o.o.	Croatia
Faculty of Food Technology and Biotechnology - Food Control Center	Croatia
Laboratory of Food and Biotechnology Institute of Agriculture and Tourism Porec	Croatia
Nastavni Zavod Za Javno Zdravstvo	Croatia
Public Health Institute	Croatia
Quality Control Laboratory	Croatia
C.P. Foodlab Ltd	Cyprus
Agricultural Research Centre	Estonia
Central Laboratory of Chemistry of Health Protection Inspectorate	Estonia
Estonian University of Life Sciences	Estonia
Laboratory of Werol Tehased Ltd	Estonia
Tallinn Veterinary and Food Laboratory	Estonia
Tartu Veterinary and Food Laboratory	Estonia
Testing Centre of University of Tartu	Estonia
Department of Crop Science	Germany
Chemical Analytical Laboratories M Galanakis	Greece
Elsap Auete	Greece
EAS Kinourias – Farmers Cooperatives	Greece
he General Chemical State Laboratory	Greece
Union of Agricultural Cooperatives of Iraklio Crete	Greece
Department of Food Chemistry, Lipid Laboratory	Hungary
General Inspectorate for Consumer Protection Laboratory of Foodstuffs and Chemicals	Hungary
National Food Investigation Institute Analytical Division Labor	Hungary
Latvenergo, Latvijan State JointStock Company	Latvian
Giih Laboratorium Kontrolno-Analitycne	Poland
ODDZIAŁ Laboratoryjny	Poland
Labor for Fermentation Technologies and Refrigeration in Food Industry	Romania
EL SPOL. S R.O.	Slovakia
Petrol d.d. , Ljubljana - Laboratory Petrol	Slovenia
Regionálny úrad verejného zdravotníctva so sídlom v Prešove, RN	Slovakia
Zavod Za Zdravstveno Varstvo Maribor-Institut Za Varstvo Okolja	Slovenia
Vitsan Gösetim Mümessilik Ve Tic. A.S.	Turkey

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[e]			Free Fatty Acid Content		
Lab number[a]	Result[b] ppm[d]	Uncertainty ppm	k[c]	Result[b] %	Uncertainty %	k[c]
1				0.0180	0.0004	2
2						
3						
4				0.03		
5				0.035	0.014	2
6	80.1	32	2	0.022	0.006	2
7				0.03	0.01	2
8	78.0	11	95%			
9	70.1	8.86	2			
10	*100*			0.055		
11	40	20	2	0.02		
12	**400**			0.022	0.006	2
13	*430*	50	95%	**0.0032**	0.00032	95%
14				0.0203	0.008	2
15				0.024		
16				0.019	0.00029	2
17				**0.26**	0.01	2
18				0.04	0.004	?
19				**0.12835**	0.001128	2
20				0.03	-	-
21				0.025	0.01	95%
22	*890*	100	2	0.057	0.003	2
23						
24	101.15	0.64	2	**0.1058**	0.03	2
25				0.026	0.002	2
26				0.039	0.0024	2.78
27						
28	102	17	2	0.028	0.003	2
29				0.065	0.0128	2
30				**60.8515**	2.43	
31	*253.8*	3	2			
32	*145.3*	348.0	2	0.0275	0.002	2
33	100	20	2	0.027	0.004	2
34						
Consensus value	**81.6**			**0.031**		

Peroxide value				Phosphorus Content in crude rapeseed oil		
Lab number[a]	Result[b] meqO2/kg	Uncertainty meqO2/kg	k[c]	Result[b] ppm[d]	Uncertainty ppm	k[c]
1	13.763	0.0306	2			
2						
3						
4	16.3	1.19	2			
5	19.93	1.6	2			
6	12.4	6.8	2	138	17	2
7	21.8	6.5	2	133	13	2
8						
9						
10	**1.12**			43.82		
11	14.6			168		
12	19.7	0.9	2			
13	14.3	1.4	95%			
14	**31.57**	4.287	2			
15	13.34					
16	7.42	0.25	2			
17	13.2	1.50	2			
18	13.7	0.068	?	125	8.74	
19	13.99	0.58	2			
20	14.4			**500**		
21	13.99	1.5	95%			
22	7.00	0.04	2			
23	14.9	0.15	2			
24	6.745	0.1	2	28.69	0.1	2
25	13.1	1.15	2	200.7	116.2	2
26						
27	15.5	2.48	2.5			
28	12.46	1.25	2	152	33	2
29	16.89	1.449	2	138.21	2.0386	2
30						
31						
32	14					
33	14.3	0.2	2			
34						
Consensus value	**14.07**			**125.3**		

	Saponification value			Beta-Sitosterol content		
Lab number[a]	Result[b] mg/g	Uncertainty mg/g	k[c]	Result[b] ppm[d]	Uncertainty ppm	k[c]
1						
2				3934	559	95%
3				3944		
4	187					
5						
6	191.5	4.8	2			
7	191	2	2			
8						
9						
10	193.5					
11						
12				3930		
13	190.1	2.1	95%			
14				3126.96	339.398	2
15						
16	190	10	2	4110	617	2
17				**1256.23**	85	2
18	192	2.51	?			
19	**204.21**	0.3	2			
20	190					
21				**1183.1786**	86	95
22						
23						
24	186.41	0.8	2			
25						
26	190.3	1.3	2			
27						
28	187	19	2			
29						
30						
31				3646.1	48	2
32	187.2					
33				3711	186	2
34				4616.85		
Consensus value	**189.7**			**3877.4**		

[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*.

One laboratory were found outlying in moisture content measurement, five laboratories in FFA measurement, two in PV measurement, one in P content measurement, one in SAPV measurement and two in B-SITO measurement (see Table 4) according to the Grubbs tests.

Peroxide value (PV) is a special parameter because of its instability. Due to some difficulties in logistics (several parcels were lost in the postal system and were resent) PV determinations were done during a long time period. Nevertheless, Figure 4 clearly demonstrates absence of any temporal trend in the PV values obtained by different laboratories at different times. Therefore all participant results were included without any special treatment.

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

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

Moisture Content	FFA content	PV	P content	SAPV	B-SITO content	
ppm	%	meq/kg	ppm	mg/g	ppm	
			EstOil-3			
81.6	0.031	14.1	125.3	189.7	3877.4	consensus value
22.4	0.013	3.6	55.4	2.3	423.5	target standard deviation
2.9	–	0.2	2.7	1.0	59.1	between-sample variability[a]
27%	42%	26%	44%	1.2%	11%	relative target standard deviation
3.6%	–	1.4%	2.2%	0.5%	1.5%	relative between-sample variability
			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.

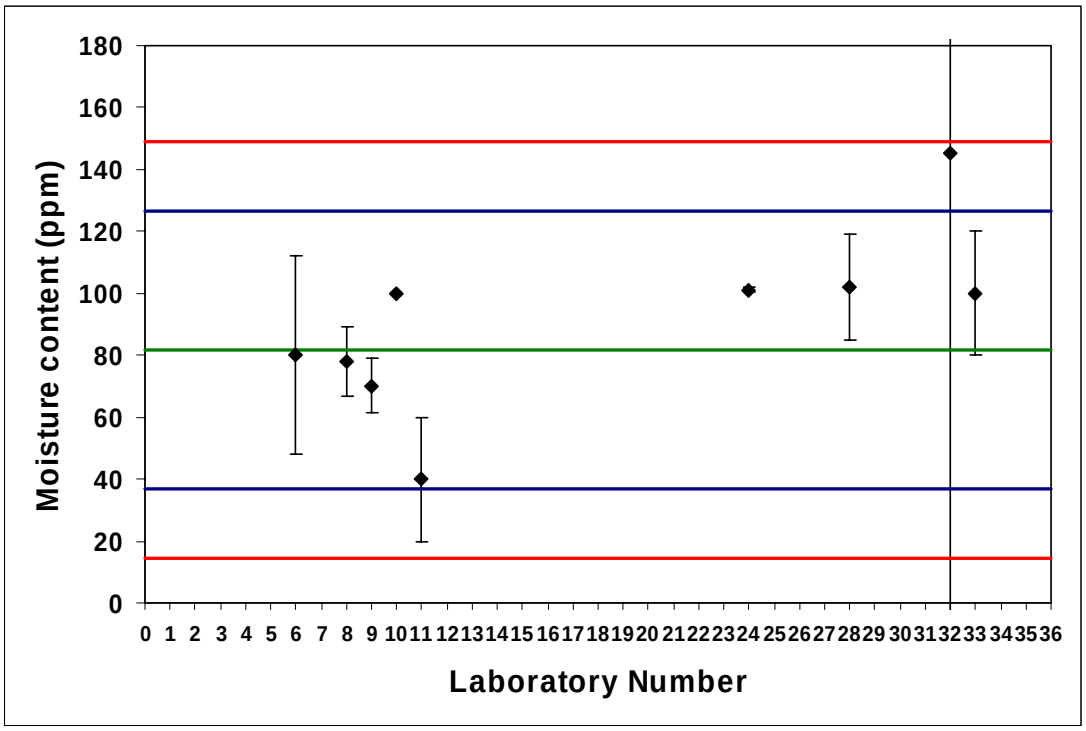
Table 6. Participant z-Scores.

Lab number[a]	z scores[b] Moisture content	FFA content	PV	P content	SAPV	B-SITO content
1		-1.0	-0.1			
2						0.1
3						0.2
4		-0.1	0.6		-1.2	
5		0.3	1.6			
6	-0.1	-0.7	-0.5	0.2	0.8	
7		-0.1	2.1	0.1	0.6	
8	-0.2					
9	-0.5					
10	0.8	1.8	-3.6	-1.5	1.7	
11	-1.9	-0.9	0.1	0.8		
12	14.2	-0.7	1.6			0.1
13	15.5	-2.2	0.1		0.2	
14		-0.8	4.9			-1.8
15		-0.6	-0.2			
16		-0.9	-1.8		0.1	0.5
17		17.5	-0.2			-6.2
18		0.7	-0.1	0.0	1.0	
19		7.4	0.0		6.4	
20		-0.1	0.1	6.8	0.1	
21		-0.5	0.0			-6.4
22	36.1	2.0	-2.0			
23			0.2			
24	0.9	5.7	-2.0	-1.7	-1.4	
25		-0.4	-0.3	1.4		
26		0.6			0.3	
27			0.4			
28	0.9	-0.3	-0.4	0.5	-1.2	
29		2.6	0.8	0.2		
30		4646.9				
31	7.7					-0.5
32	2.8	-0.3	0.0		-1.1	
33	0.8	-0.3	0.1			-0.4
34						1.7

[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.

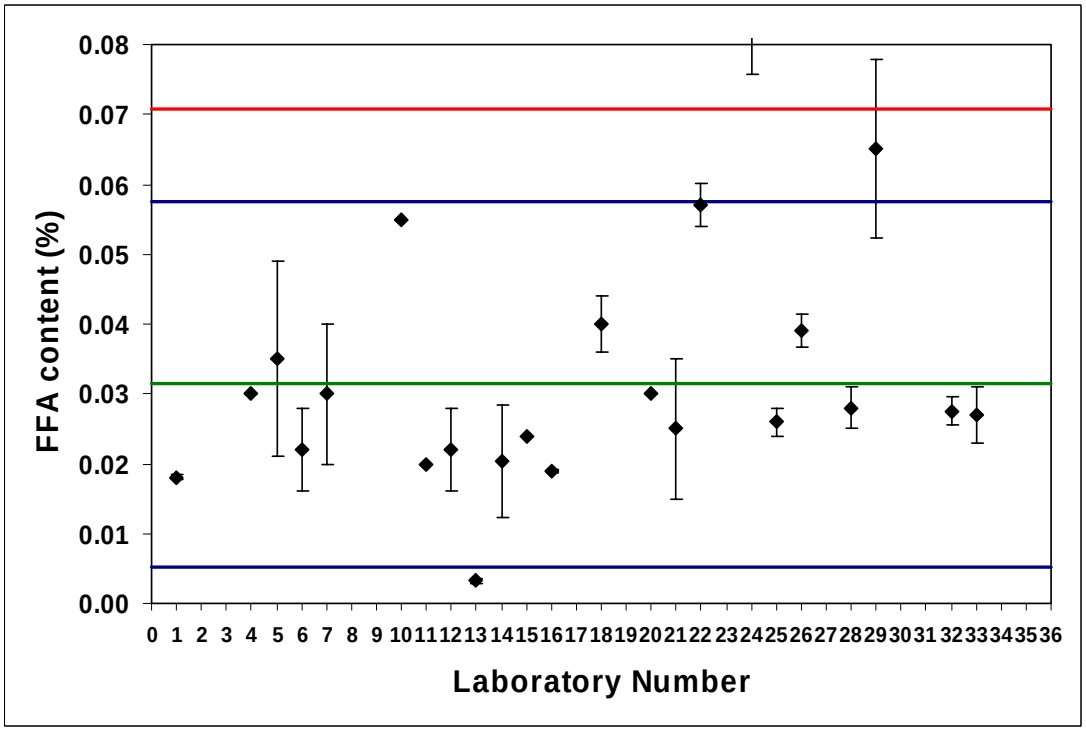
The results are presented in graphical form in the Figures below:

Figure 1. Results of Participants with the z-Score Boundaries.[a] Moisture Content Measurement[b].



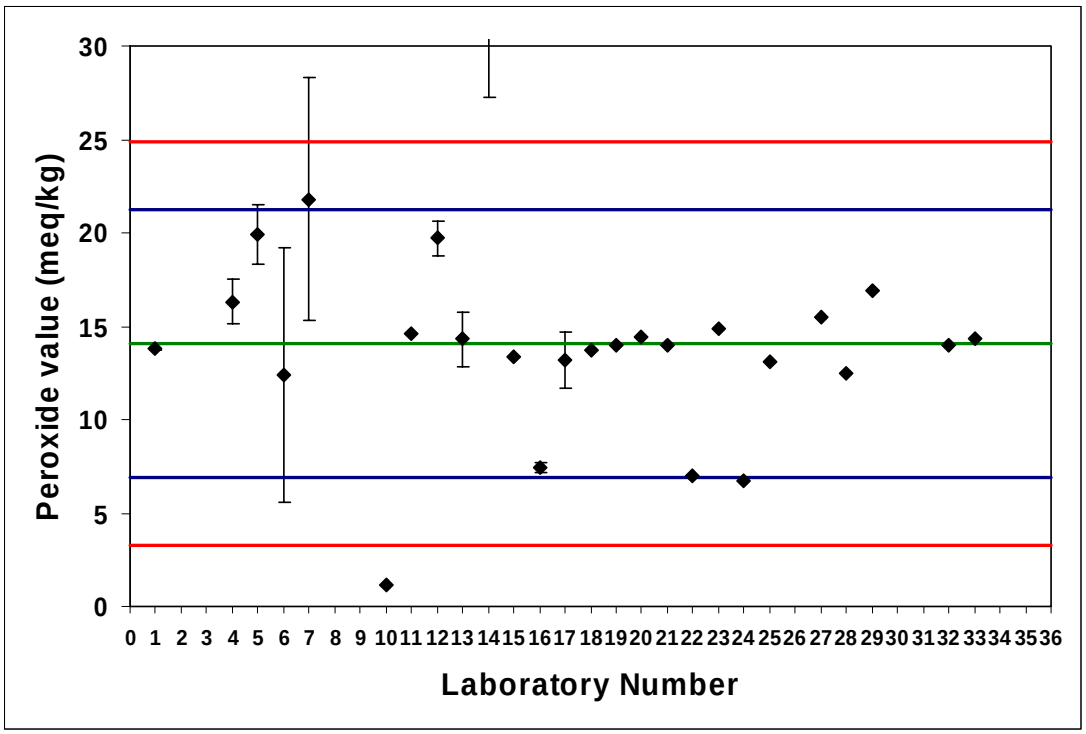
[a] The consensus value is denoted by a green line. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b] Results of the laboratories 12, 13, 22 and 31 are out of the figure range (see the Table 4).

Figure 2. Results of Participants with the z-Score Boundaries.[a] FFA Content Measurement[b].



[a] The consensus value is denoted by a green line. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b] Results of the laboratories 17, 19, 24 and 30 are out of the figure range (see the Table 4).

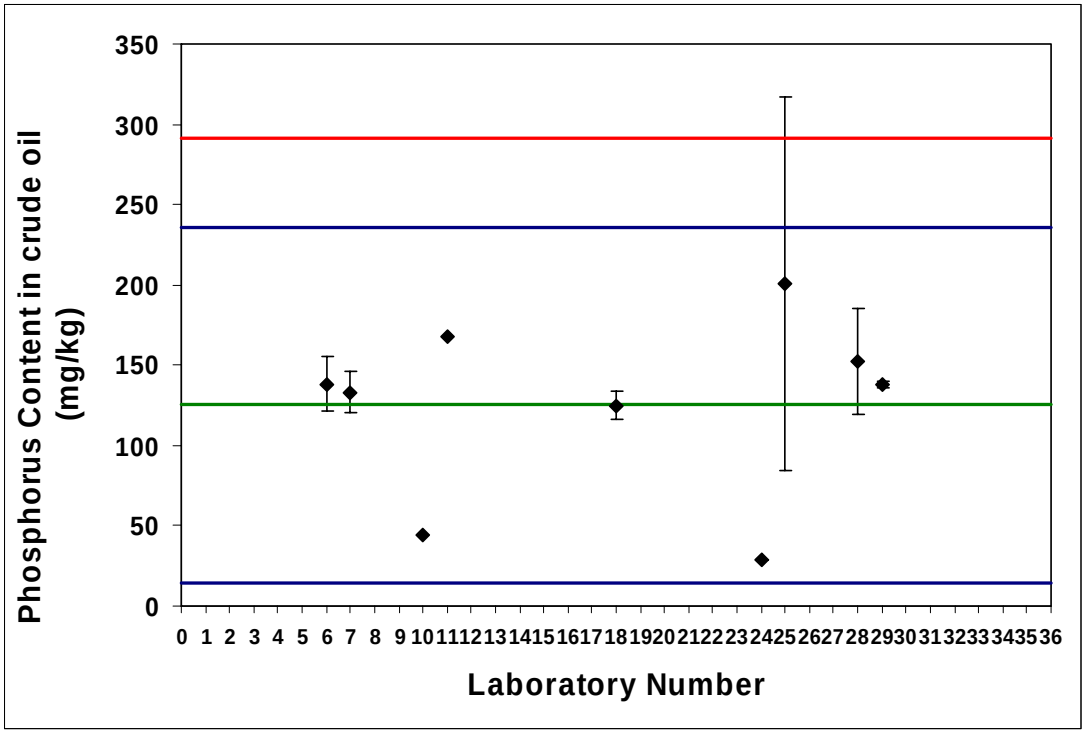
Figure 3. Results of Participants with the z-Score Boundaries.[a] PV Measurement[b].



[a] The consensus value is denoted by a green line. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b] Result of the laboratory 14 is out of the figure range (see the Table 4).

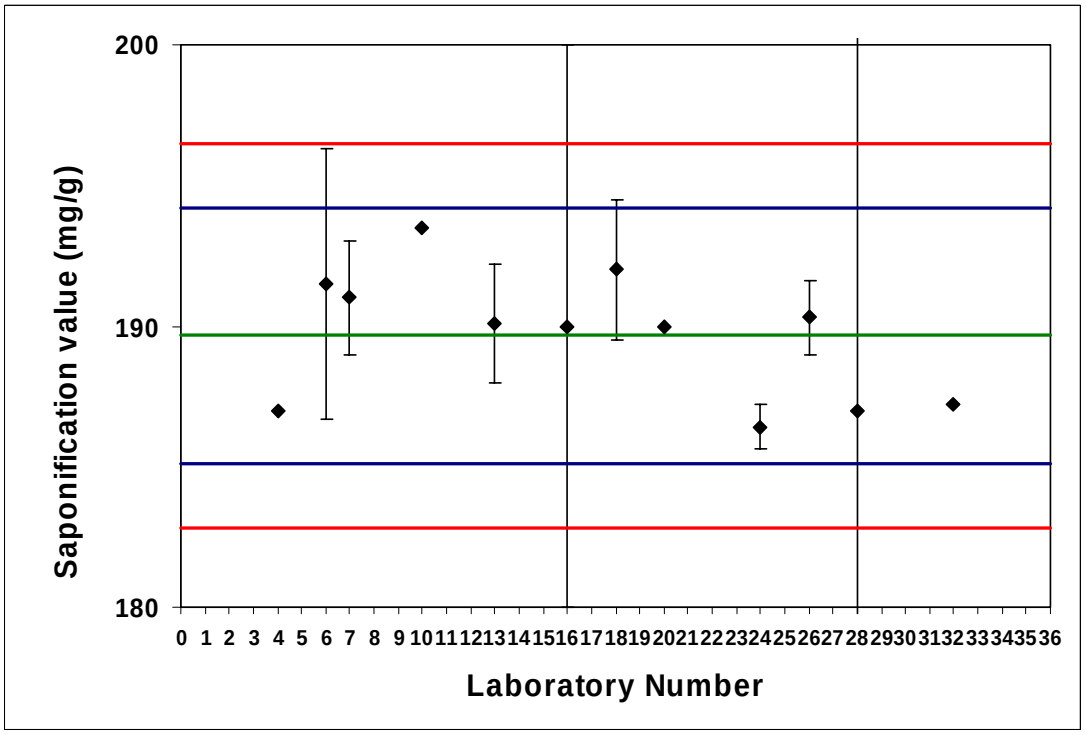
Figure 4. Dependence of the Results of PV Measurement on Measurement Date.

Figure 5. Results of Participants with the z-Score Boundaries.[a] P Content Measurement[b].



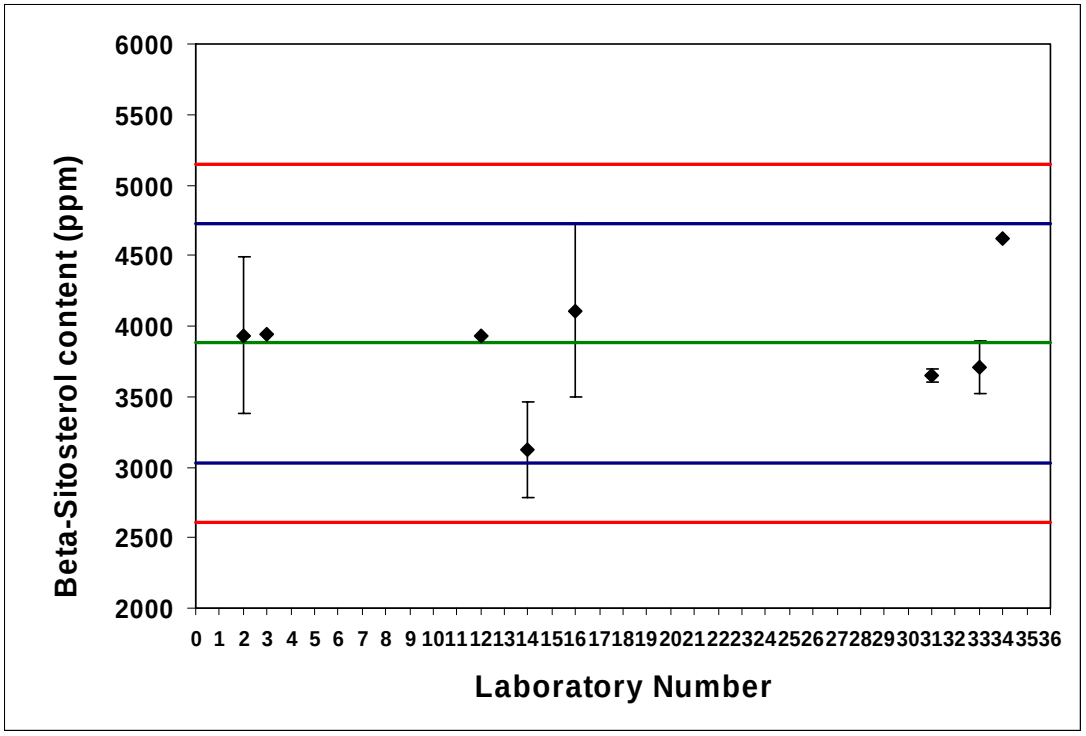
[a] The consensus value is denoted by a green line. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b] Result of the laboratory 20 is out of the figure range (see the Table 4).

Figure 6. Results of Participants with the z-Score Boundaries.[a] SAPV Measurement[b].



[a] The consensus value is denoted by a green line. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b] Result of the laboratory 19 is out of the figure range (see the Table 4).

Figure 7. Results of Participants with the z-Score Boundaries.[a] B-SITO Measurement[b].



[a] The consensus value is denoted by a green line. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b] Results of the laboratories 17 and 21 are out of the figure range (see the Table 4).

From Table 6 and the Figures it can be concluded that based on the z-score approach from the 98 submitted results 14 are unacceptable and 4 are doubtful.

4.2 Between-Sample Variability

Between-sample variability was determined by UT (moisture content), by WT (FFA content, PV, P content, SAPV) and by Estonian University of Life Sciences (B-SITO 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.[6]

For the moisture content the between-sample standard deviation found by UT was 2.9 ppm (five samples). This is 8 times lower than the between-participant standard deviation. For the FFA content the between-sample standard deviation could not be calculated because the ANOVA results indicated that within-sample variability is larger than the overall variability. This is of course impossible: this result is due to statistical fluctuations in data and indicates that between-sample variability is negligible. For the PV the between-sample standard deviation was found 0.2 meqO2/kg (four bottles), which is 18 times lower than the between-participant standard deviation. For the phosphorus content the between-sample standard deviation was found 2.7 ppm (four

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

bottles), which is 21 times lower than the between-participant standard deviation. For the SAPV the between-sample standard deviation was found 1.0 mg/g (four bottles), which is 2 times lower than the between-participant standard deviation. For the B-SITO the between-sample standard deviation was found 59 ppm, which is 7 times lower than the between-participant standard deviation.

We can conclude that in moisture content, FFA content, PV, P content and B-SITO content measurement the between-sample variability has negligible effect on the between-participant variability. In the case of SAPV measurement the difference is only 2 times and this between-sample variability is not negligible anymore. Therefore with this parameter the between-sample variability is taken into account in E_n score calculation (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 (see section 5.2 for discussion) and not by sample instability.

5 Discussion

5.1 Elimination of Outliers Using the Two-Step Grubbs Test

Altogether 12 participant results were eliminated from consensus value and target standard deviation calculation based on the Grubbs tests. Out of these two were rejected using the second Grubbs test with logarithmic values (see section 4.1). Both of these two values were around ten times lower than the respective consensus values and around six times lower than the next lowest values. Based on this we find that the two-step approach is well justified in cases when the spread of the participant results is high.

5.2 Assessment of Participant Results by the z-Score Approach

All in all 98 results were submitted. According to the z-score approach 80 of them (81.6%) were acceptable, 4 (4.1%) were doubtful and 14 (14.3%) were unacceptable (See Table 6).

At first sight it may seem that the results are good, but closer inspection of the data reveals that with most of the measurands the picture is far less satisfactory than it seems at first.

The large number of acceptable z-scores is primarily caused by using the actual standard deviations as target standard deviations in z-score calculation. Even after outlier elimination the standard deviations of moisture, FFA, PV and P are still large being 25 .. 45% of the consensus value. This causes the width of the acceptable result zone (± 2*s*) to be 1.0 to 1.8 times the consensus value! The situation with beta-sitosterol content (relative target standard deviation 11%) is a lot better. The situation with saponification value is very good – the relative target standard deviation is 1.2% of the consensus value.

The spread of the participant results in beta-sitosterol determination is good and in saponification value determination excellent.

This very low spread of the participant values also explains why the between-sample variability in saponification value determination is only two times 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 laboratory biases. These could be due to the following factors:

1. The contents of moisture and FFA in the samples were low. This is probably one of the main reasons for the large spread of FFA content measurement (measuring near of LoQ). With moisture and FFA both consensus values and the target standard deviations are similar to those of EstOil-2.

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

2. PV is an unstable measurand. In spite of this, the spread of PV measurement results is quite reasonable being the lowest of the four problematic measurands.
3. Different methods were used by different participants. This can have an effect on the agreement of the participant results. In particular, based on the reported participant results and on the results of our own study, we conclude that the heating procedure (based on the ISO 662 standard) is unsuitable because of the low moisture content of the samples. This is at variance with our findings in EstOil-2. However, there was just a single result reported with loss on heating in EstOil-2. Apart from this, we did not observe any trends or systematic differences between results of laboratories comparing these methods. The results of participants using identical methods often differed very much.
4. Sample-to-sample variability plays insignificant role (see previous section), with 5 out of 6 parameters (the exception is the saponification value).

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 results. It can easily be seen that

The majority of the participants have underestimated their uncertainties in moisture, FFA and PV determination.

For example, in FFA content measurements the uncertainty intervals of laboratories 1, 22, 26 and 28 do not overlap in any of the two-lab pairs! This means that at least three of the four laboratories have underestimated the uncertainties of their results. At the same time all these laboratories have satisfactory z-score values.

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	6	8	9	11	13	22	24	28	31	32	33
6	0.0	0.1	0.3	1.1	5.9	7.7	0.7	0.6	5.4	0.2	0.5
8	0.1	0.0	0.6	1.7	6.9	8.1	2.1	1.2	15.4	0.2	1.0
9	0.3	0.6	0.0	1.4	7.1	8.2	3.5	1.7	19.6	0.2	1.4
11	1.1	1.7	1.4	0.0	7.2	8.3	3.1	2.4	10.6	0.3	2.1
13	5.9	6.9	7.1	7.2	0.0	4.1	6.6	6.2	3.5	0.8	6.1
22	7.7	8.1	8.2	8.3	4.1	0.0	7.9	7.8	6.4	2.1	7.7
24	0.7	2.1	3.5	3.1	6.6	7.9	0.0	0.0	49.1	0.1	0.1
28	0.6	1.2	1.7	2.4	6.2	7.8	0.0	0.0	8.8	0.1	0.1
31	5.4	15.4	19.6	10.6	3.5	6.4	49.1	8.8	0.0	0.3	7.6
32	0.2	0.2	0.2	0.3	0.8	2.1	0.1	0.1	0.3	0.0	0.1
33	0.5	1.0	1.4	2.1	6.1	7.7	0.1	0.1	7.6	0.1	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.

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

	$\lvert E_n \rvert$ value																				
Lab No	1	5	6	7	12	13	14	16	17	18	19	21	22	24	25	26	28	29	30	32	33
1	0.0	1.2	0.7	1.2	0.7	30.7	0.3	2.2	24.2	5.5	93.2	0.7	12.9	2.9	3.9	8.7	3.3	3.7	25.0	4.7	2.2
5	1.2	0.0	0.9	0.3	0.9	2.3	0.9	1.1	13.1	0.3	6.6	0.6	1.5	2.1	0.6	0.3	0.5	1.6	25.0	0.5	0.5
6	0.7	0.9	0.0	0.7	0.0	3.1	0.2	0.5	20.4	2.5	17.4	0.3	5.2	2.7	0.6	2.6	0.9	3.0	25.0	0.9	0.7
7	1.2	0.3	0.7	0.0	0.7	2.7	0.8	1.1	16.3	0.9	9.8	0.4	2.6	2.4	0.4	0.9	0.2	2.2	25.0	0.2	0.3
12	0.7	0.9	0.0	0.7	0.0	3.1	0.2	0.5	20.4	2.5	17.4	0.3	5.2	2.7	0.6	2.6	0.9	3.0	25.0	0.9	0.7
13	30.7	2.3	3.1	2.7	3.1	0.0	2.1	36.9	25.7	9.2	106.7	2.2	17.8	3.4	11.3	14.8	8.2	4.8	25.0	12.0	5.9
14	0.3	0.9	0.2	0.8	0.2	2.1	0.0	0.2	18.7	2.2	13.4	0.4	4.3	2.8	0.7	2.2	0.9	3.0	25.0	0.9	0.7
16	2.2	1.1	0.5	1.1	0.5	36.9	0.2	0.0	24.1	5.2	94.0	0.6	12.6	2.9	3.5	8.3	3.0	3.6	25.0	4.2	2.0
17	24.2	13.1	20.4	16.3	20.4	25.7	18.7	24.1	0.0	20.4	13.1	16.6	19.4	4.9	22.9	21.5	22.2	12.0	24.9	22.8	21.6
18	5.5	0.3	2.5	0.9	2.5	9.2	2.2	5.2	20.4	0.0	21.3	1.4	3.4	2.2	3.1	0.2	2.4	1.9	25.0	2.8	2.3
19	93.2	6.6	17.4	9.8	17.4	106.7	13.4	94.0	13.1	21.3	0.0	10.3	22.3	0.8	44.6	33.7	31.3	4.9	25.0	43.9	24.1
21	0.7	0.6	0.3	0.4	0.3	2.2	0.4	0.6	16.6	1.4	10.3	0.0	3.1	2.6	0.1	1.4	0.3	2.5	25.0	0.2	0.2
22	12.9	1.5	5.2	2.6	5.2	17.8	4.3	12.6	19.4	3.4	22.3	3.1	0.0	1.6	8.6	4.7	6.8	0.6	25.0	8.2	6.0
24	2.9	2.1	2.7	2.4	2.7	3.4	2.8	2.9	4.9	2.2	0.8	2.6	1.6	0.0	2.7	2.2	2.6	1.3	25.0	2.6	2.6
25	3.9	0.6	0.6	0.4	0.6	11.3	0.7	3.5	22.9	3.1	44.6	0.1	8.6	2.7	0.0	4.2	0.6	3.0	25.0	0.5	0.2
26	8.7	0.3	2.6	0.9	2.6	14.8	2.2	8.3	21.5	0.2	33.7	1.4	4.7	2.2	4.2	0.0	2.9	2.0	25.0	3.7	2.5
28	3.3	0.5	0.9	0.2	0.9	8.2	0.9	3.0	22.2	2.4	31.3	0.3	6.8	2.6	0.6	2.9	0.0	2.8	25.0	0.1	0.2
29	3.7	1.6	3.0	2.2	3.0	4.8	3.0	3.6	12.0	1.9	4.9	2.5	0.6	1.3	3.0	2.0	2.8	0.0	25.0	2.9	2.8
30	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	24.9	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	0.0	25.0	25.0
32	4.7	0.5	0.9	0.2	0.9	12.0	0.9	4.2	22.8	2.8	43.9	0.2	8.2	2.6	0.5	3.7	0.1	2.9	25.0	0.0	0.1
33	2.2	0.5	0.7	0.3	0.7	5.9	0.7	2.0	21.6	2.3	24.1	0.2	6.0	2.6	0.2	2.5	0.2	2.8	25.0	0.1	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.

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

	$\lvert E_n \rvert$ value																				
Lab No	1	4	5	6	7	12	13	14	16	17	18	19	21	22	23	24	25	27	28	29	33
1	0.0	2.1	3.9	0.2	1.2	6.6	0.4	4.2	25.2	0.4	0.8	0.4	0.2	134.3	7.5	67.1	0.6	0.7	1.0	2.2	2.5
4	2.1	0.0	1.8	0.6	0.8	2.3	1.1	3.4	7.3	1.6	2.2	1.7	1.2	7.8	1.2	8.0	1.9	0.3	2.2	0.3	1.7
5	3.9	1.8	0.0	1.1	0.3	0.1	2.6	2.5	7.7	3.1	3.9	3.5	2.7	8.1	3.1	8.2	3.5	1.5	3.7	1.4	3.5
6	0.2	0.6	1.1	0.0	1.0	1.1	0.3	2.4	0.7	0.1	0.2	0.2	0.2	0.8	0.4	0.8	0.1	0.4	0.0	0.6	0.3
7	1.2	0.8	0.3	1.0	0.0	0.3	1.1	1.3	2.2	1.3	1.2	1.2	1.2	2.3	1.1	2.3	1.3	0.9	1.4	0.7	1.2
12	6.6	2.3	0.1	1.1	0.3	0.0	3.2	2.7	13.1	3.7	6.6	5.3	3.3	14.1	5.3	14.3	4.5	1.6	4.7	1.6	5.8
13	0.4	1.1	2.6	0.3	1.1	3.2	0.0	3.8	4.7	0.5	0.4	0.2	0.1	5.1	0.4	5.3	0.7	0.4	1.0	1.3	0.0
14	4.2	3.4	2.5	2.4	1.3	2.7	3.8	0.0	5.6	4.0	4.2	4.1	3.9	5.7	3.9	5.8	4.2	3.2	4.3	3.2	4.0
16	25.2	7.3	7.7	0.7	2.2	13.1	4.7	5.6	0.0	3.8	24.2	10.4	4.3	1.7	25.7	2.5	4.8	3.2	4.0	6.4	20.9
17	0.4	1.6	3.1	0.1	1.3	3.7	0.5	4.0	3.8	0.0	0.3	0.5	0.4	4.1	1.1	4.3	0.1	0.8	0.4	1.8	0.7
18	0.8	2.2	3.9	0.2	1.2	6.6	0.4	4.2	24.2	0.3	0.0	0.5	0.2	84.9	7.3	57.5	0.5	0.7	1.0	2.2	2.7
19	0.4	1.7	3.5	0.2	1.2	5.3	0.2	4.1	10.4	0.5	0.5	0.0	0.0	12.0	1.5	12.3	0.7	0.6	1.1	1.9	0.5
21	0.2	1.2	2.7	0.2	1.2	3.3	0.1	3.9	4.3	0.4	0.2	0.0	0.0	4.7	0.6	4.8	0.5	0.5	0.8	1.4	0.2
22	134.3	7.8	8.1	0.8	2.3	14.1	5.1	5.7	1.7	4.1	84.9	12.0	4.7	0.0	51.2	2.4	5.3	3.4	4.4	6.8	33.5
23	7.5	1.2	3.1	0.4	1.1	5.3	0.4	3.9	25.7	1.1	7.3	1.5	0.6	51.2	0.0	45.4	1.6	0.2	1.9	1.4	2.3
24	67.1	8.0	8.2	0.8	2.3	14.3	5.3	5.8	2.5	4.3	57.5	12.3	4.8	2.4	45.4	0.0	5.5	3.5	4.6	7.0	31.9
25	0.6	1.9	3.5	0.1	1.3	4.5	0.7	4.2	4.8	0.1	0.5	0.7	0.5	5.3	1.6	5.5	0.0	0.9	0.4	2.0	1.0
27	0.7	0.3	1.5	0.4	0.9	1.6	0.4	3.2	3.2	0.8	0.7	0.6	0.5	3.4	0.2	3.5	0.9	0.0	1.1	0.5	0.5
28	1.0	2.2	3.7	0.0	1.4	4.7	1.0	4.3	4.0	0.4	1.0	1.1	0.8	4.4	1.9	4.6	0.4	1.1	0.0	2.3	1.5
29	2.2	0.3	1.4	0.6	0.7	1.6	1.3	3.2	6.4	1.8	2.2	1.9	1.4	6.8	1.4	7.0	2.0	0.5	2.3	0.0	1.8
33	2.5	1.7	3.5	0.3	1.2	5.8	0.0	4.0	20.9	0.7	2.7	0.5	0.2	33.5	2.3	31.9	1.0	0.5	1.5	1.8	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	6	7	18	24	25	28	29
6	0.0	0.2	0.7	6.4	0.5	0.4	0.0
7	0.2	0.0	0.5	8.0	0.6	0.5	0.4
18	0.7	0.5	0.0	11.0	0.6	0.8	1.5
24	6.4	8.0	11.0	0.0	1.5	3.7	53.7
25	0.5	0.6	0.6	1.5	0.0	0.4	0.5
28	0.4	0.5	0.8	3.7	0.4	0.0	0.4
29	0.0	0.4	1.5	53.7	0.5	0.4	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.

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

	$\lvert E_n \rvert$ value								
Lab No	**6**	**7**	**13**	**16**	**18**	**19**	**24**	**26**	**28**
6	0	0.1	0.2	0.1	0.1	2.3	0.9	0.2	0.2
7	0.1	0	0.2	0.1	0.2	3.8	1.3	0.2	0.2
13	0.2	0.2	0	0.0	0.4	4.0	1.0	0.1	0.2
16	0.1	0.1	0.0	0	0.2	1.4	0.3	0.0	0.1
18	0.1	0.2	0.4	0.2	0	3.3	1.5	0.4	0.3
19	2.3	3.8	4.0	1.4	3.3	0	6.1	4.5	0.9
24	0.9	1.3	1.0	0.3	1.5	6.1	0	1.2	0.0
26	0.2	0.2	0.1	0.0	0.4	4.5	1.2	0	0.2
28	0.2	0.2	0.2	0.1	0.3	0.9	0.0	0.2	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.

	$\lvert E_n \rvert$ value						
Lab No	**2**	**14**	**16**	**17**	**21**	**31**	**33**
2	0.0	1.2	0.2	4.7	4.9	0.5	0.4
14	1.2	0.0	1.4	5.3	5.6	1.5	1.5
16	0.2	1.4	0.0	4.6	4.7	0.8	0.6
17	4.7	5.3	4.6	0.0	0.6	24.4	12.0
21	4.9	5.6	4.7	0.6	0.0	24.9	12.4
31	0.5	1.5	0.8	24.4	24.9	0.0	0.3
33	0.4	1.5	0.6	12.0	12.4	0.3	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 the Tables it can be seen that in phosphorus and saponification value determination between-lab agreements dominate. With both of these measurands there is basically only one laboratory (24 and 19, respectively) whose result is in disagreement with others.

The situation is distinctly different with the remaining four measurands. Disagreeing comparisons dominate: moisture 65%, FFA 71%, PV 68% and beta-sitosterol 67% of the comparison pairs have disagreement. Clearly, most of the participants should take a close look at their uncertainty estimates of moisture content, FFA, PV and also beta-sitosterol content measurements. The abovementioned factors provide some guidelines. As a conclusion:

The pair-wise agreement of participant results in phosphorus and saponification value determination is satisfactory while in the determination of the remaining four parameters it is unsatisfactory.

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

6 Acknowledgments

The organizers would like to thank Mr Viktor Vabson (AS Metrosert, Division of Standards Services), Mr. Rouvim Kadis (D. I. Mendeleyev Institute for Metrology, Russia), Dr. Koit Herodes (University of Tartu, Department of Chemistry) and MSc Erkki Mäeorg (Estonian University of Life Sciences, Institute of Agricultural and Environmental Sciences) for advice and help.

7 Annex 1

The usual statistical algorithm of finding z scores[3] 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.[5] 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)

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

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 13.

Table 13. The median-based consensus values for the measurands. Mean-based consensus values are given for comparison.

Moisture content	FFA content	PV	P content	SAPV	B-SITO content	
ppm	%	meq/kg	ppm	mg/g	ppm	unit
			EstOil-3			
81.6	0.03	14.1	125.3	189.7	3877.4	consensus value
101.6	0.03	14.0	138.1	190.1	3820.5	consensus value (median)
22.4	0.01	3.6	55.4	2.3	423.5	target standard deviation
16.7	0.003	0.4	13.6	1.0	143.7	target standard deviation (median)

2.9	0.00	0.2	2.7	1.0	59.1	between-sample variability
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

Assessment of the results is done using the E_n numbers as described in ISO Guide 43-1:[3]

$$|E_n| = \frac{|C_{lab} - C_c|}{\sqrt{U_{lab}^2 + U_c^2}} \quad (9)$$

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.

	$\|E_n\|$ numbers[b] accdording to median approach					
Lab number[a]	Moisture content	FFA	PV	P content	SAPV	B-SITO content
1		1.7	0.3			
2						0.2
3						0.4
4		0.2	1.5		1.5	
5		0.4	3.2			
6	0.5	0.8	0.2	0.0	0.3	
7		0.1	1.2	0.2	0.3	
8	0.7					
9	0.9					
10	0.0	4.0	14.4	3.5	1.7	
11	1.6	1.4	0.7	1.1		
12	9.0	0.8	4.5			0.4
13	5.5	3.9	0.2		0.0	
14		0.8	4.0			1.6
15		0.8	0.7			
16		1.5	7.1		0.0	0.4
17		19.3	0.5			8.6
18		1.4	0.3	0.5	0.6	
19		14.9	0.0		6.8	
20		0.2	0.5	13.3	0.0	
21		0.3	0.0			8.8
22	7.5	3.9	7.8			

23			1.0			
24	0.0	2.5	8.0	4.0	1.7	
25		0.4	0.6	0.5		
26		1.4			0.1	
27			0.6			
28	0.0	0.1	1.0	0.3	0.2	
29		2.5	1.7	0.0		
30		25.0				
31	4.6					0.6
32	0.1	0.2	0.0		1.4	
33	0.0	0.3	0.3			0.3
34						2.8

[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 or of informative nature for the current intercomparison round but more investigations will be performed and in the future this data treatment will be considered as the definitive one. If this decision will be made then this will be stated in the invitation to the intercomparison.