Interlaboratory Comparison Measurement EstOil-1

Final report

Final Version 25.07.2005

Report Compiled by

Ivo Leito

Tiina Kukk

Testing Centre, University of Tartu

Werol Tehased AS

Tartu 2005

Table of Contents

1 The Aim of the Intercomparison 3

2 Organization 3
2.1 General 3
2.2 The Samples 3
2.3 Data Treatment 3

3 Participants 4

4 Results 5
4.1 Results of the Participants 5
4.2 Between-Sample Variability 7

5 Discussion 8

1 The Aim of the Intercomparison

The aim of the EstOil-1 intercomparison was to allow the participating laboratories to assess their performance in determining two edible oil parameters: moisture content (by Karl Fischer titration) and free fatty acids (FFA) content.

2 Organization

2.1 General

The intercomparison measurement was organized jointly by University of Tartu Testing Centre (below UT) and Werol Tehased AS (below WT).

The Final report was compiled jointly by UT and WT and will be made public (on the website of UT). The participants are listed in the final 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 samples of approximately 120 ml in sealed glass ampoules. The samples were prepared from a single bulk of oil that was well mixed before sealing into ampoules. The laboratories got random ampoules from the pool of ampoules. The first and last ampoule were not distributed.

2.3 Data Treatment

The evaluation of participant data was done at UT according to the ISO Guide 43-1.[1] 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 possible outlier results). The target standard deviation in our case is found as the real standard deviation of the participant values (after elimination of possible outliers).

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

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

Table 1. Assessemnt 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

3 Participants

Invitations were sent to a number of laboratories in Estonia and other countries from Eastern Europe. The participants are listed in Table 2.

Table 2. Participants to EstOil-1.

Institution	Country
University of Tartu, testing Centre	Estonia
Werol Tehased AS	Estonia
Areto OÜ	Estonia
Central Chemical Laboratory of Health Protection Inspectorate	Estonia
Petrol d.d., Ljubljana – Laboratory Petrol	Slovenia
Laboratory for Fermentation Technologies and Refrigeration in Food Industry	Romania
VAS Latvenergo, "High Voltage Network", IPAG Chemical Laboratory	Latvia
VAS Latvenergo, "Riga Thermal Power Plants", Chemical laboratory	Latvia
Oil Testing Laboratory Equipment Operation Division Lietuvos Energio AB	Lithuania

4 Results

4.1 Results of the Participants

Results of the participants are presented in Table 3 the same way as provided by the participants. In certain cases participants were asked to convert their values, so that uniformity of units was achieved.

Table 3. 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[b] ppm	*k*[c]	Result[b] %	Uncertainty[b] %	*k*[c]
1	400	12	2	0.04975	0.00248	2
2	328.7	3.4	2.04			
3				0.08	0.0028	2
4	381	38	2	0.067	0.006	2
5	337.0	5.5	2			
6				0.06	0.02	2
7	377.5	5	2			
8	393	19.65	2			
9	320	20	2	0.08		
Consensus value	**362.5**			**0.067**		

[a] The participating laboratories are given numbers in random order that is different from the order given in Table 2. [b] The results are presented the same way as they were presented by the participants. [c] Coverage factor. [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.

There are no outliers in any of the two measured parameters. The consensus values were derived from the participant data as average values: 362.5 ppm and 0.067 % for the moisture content and FFA content, respectively. The standard deviations of the participant data were used as the target standard deviations: 32.9 ppm and 0.013 % for the moisture content and FFA content, respectively. The z-scores are calculated according to equation 1 and are presented in Table 4.

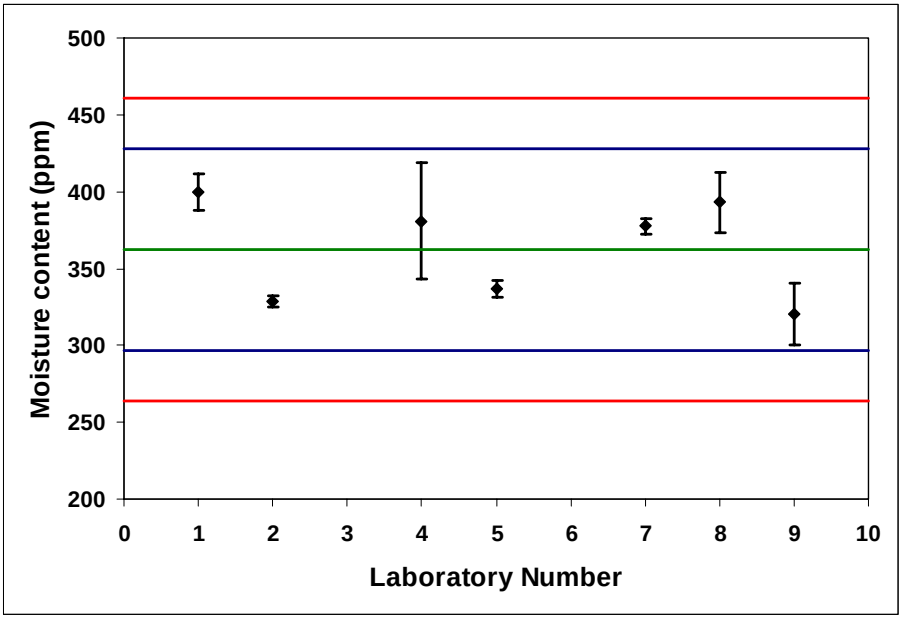
Table 4. Participant z-Scores.

Lab[a] Number	z scores Moisture content	FFA content
1	1.14	-1.35
2	-1.03	
3		0.97
4	0.56	-0.03
5	-0.77	
6		-0.56
7	0.46	
8	0.93	
9	-1.29	0.97

[a] The participating laboratories are given in random order that is different from the order given in Table 2 but is identical to the order given in Table 3.

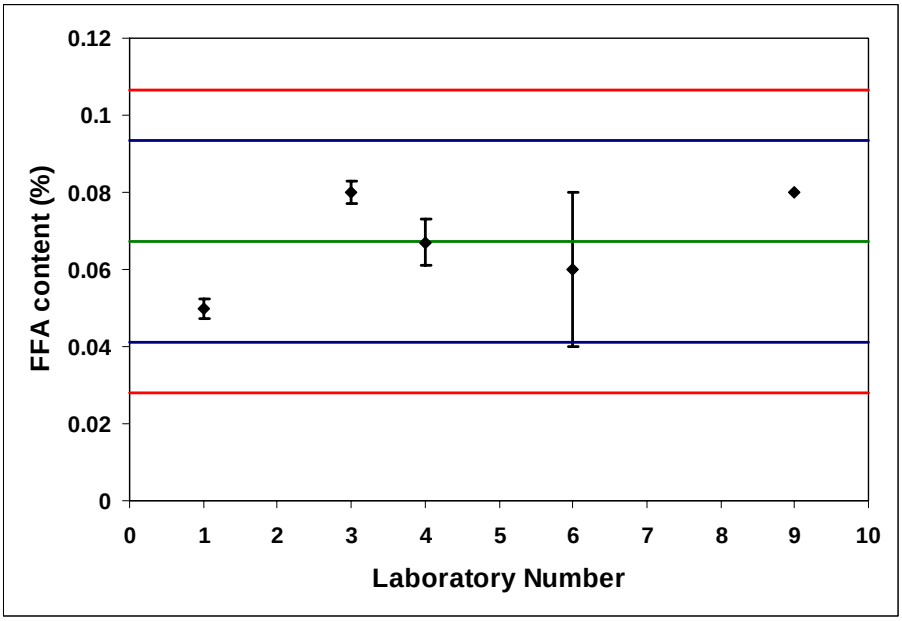
The results are presented in graphical form in the Figures below:

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



[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.

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



[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.

From Table 4 and the Figures it can be concluded that based on the z-score approach all the participating laboratories have performed satisfactorily.

4.2 Between-Sample Variability

Between-sample variability was determined by Werol Tehased AS (both parameters) and by University of Tartu (moisture content). The data were treated using the ANOVA technique to separate the effects of between- and within-sample variability.[2]

For the FFA content the between-sample variability standard deviation was found 0.00035% (four ampoules).

For the Humidity content the between-sample variability standard deviation found by Werol tehased AS was 6.0 ppm (five ampoules). For the results obtained at UT Testing Centre the ANOVA method indicated negligible difference between the ampoules (two ampoules). The difference between the means was only 0.1 ppm.

We can conclude that between-sample variability is negligible in measuring both parameters.

It is necessary to note, however, that the difference between the mean moisture contents obtained at UT Testing Centre and Werol Tehased AS was around 40 ppm indicating significant laboratory bias also between the organizers.

[2] 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

The values obtained by the participants are spread more or less homogenously. There are no outliers in either of the measurements. However, the spread of the values is significant and as indicated by the between-ampoule variability determinations, is most probably caused by large laboratory biases. These could be due to the following factors:

1. The values of both measurands in the samples were low. This is probably one of the main reasons for the large spread. If one examines for example the ISO standards for moisture content and FFA content measurement – ISO 8534 and ISO 660 respectively – then one discovers that in both cases the values of the measurands in the samples were strongly on the low side. These low values amplify the influence of the uncertainty sources.
2. Different standard methods were used by different participants. This can have an effect even though the basics of the methods are identical (Karl Fischer titration and acid-base titration) some differences introduce variability of the results. Nevertheless this is probably not the most important cause of the spread, because the results of the laboratories that used the same method, can also differ very much. As an example, laboratories 2, 7 and 8 all used the same IEC 60814 standard method for moisture content measurement and the spread of their results is 64.3 ppm, which is only slightly less than the full spread of all participant results (80 ppm).
3. Especially with volumetric Karl Fischer titration it is very important to regularly re-standardize the titrant. Deviation of the titrant's real concentration from the assumed concentration can cause a large laboratory bias.
4. Sample-to-sample variability plays probably insignificant role (see previous section), however, it cannot be ruled out. Also, several participants reported that it was difficult to open the sample ampoules and the ampoules could not be conveniently closed once they were opened. Difficulties in handling the ampoules can cause some moisture uptake from the atmosphere.

Although according to the z-score approach all 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 from such a comparison that most probably the majority of the participants have underestimated their uncertainties.** For example, in moisture content measurements the uncertainty intervals of laboratories 1, 2 and 7 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!

Very illuminating in this respect are comparisons of the results of the participants in pairs. This is done using the E_n numbers as described in ISO Guide 43-1:[1]

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

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. Agreement between two results is considered acceptable if $|E_n| \leq 1$. The paired comparisons are presented in Table 5 and Table 6.

Table 5. Comparison of the Results of Participants in Pairs: Moisture Content.

	E_n Value						
Lab No	**1**	**2**	**4**	**5**	**7**	**8**	**9**
1	0	-5.7	-0.5	-4.8	-1.7	-0.3	-3.4
2	5.7	0	1.4	1.3	8.1	3.2	-0.4
4	0.5	-1.4	0	-1.1	-0.1	0.3	-1.4
5	4.8	-1.3	1.1	0	5.4	2.7	-0.8
7	1.7	-8.1	0.1	-5.4	0	0.8	-2.8
8	0.3	-3.2	-0.3	-2.7	-0.8	0	-2.6
9	3.4	0.4	1.4	0.8	2.8	2.6	0

The numbers of participants are the same as in Table 3.

Table 6. Comparison of the Results of Participants in Pairs: Free Fatty Acid Content.

	E_n value				
Lab No	**1**	**3**	**4**	**6**	**9**
1	0	8.1	2.7	0.5	12.2
3	-8.1	0	-2.0	-1.0	0.0
4	-2.7	2.0	0	-0.3	2.2
6	-0.5	1.0	0.3	0	1.0
9	-12.2	0.0	-2.2	-1.0	0

The numbers of participants are the same as in Table 3.

As can be seen from the E_n values only 7 out of 21 pairs display satisfactory agreement in moisture content measurement and 5 out of 10 in FFA content measurement. Clearly, most of the participants should take a close look at their uncertainty estimates. The abovementioned factors provide some guidelines.

An alternative and more rigorous way to assess the performance of the participants would be to compare their results to independent (metrological) reference values of the measurands. This is, however, impossible in the current intercomparison, because the consensus values of the measurands obtained from the participant data are not trustworthy enough.

The general conclusions from the results can thus be worded as follows:

1. **Given the low values of the measured parameters in the samples the agreement of the obtained values, leaving aside their claimed uncertainties, can be considered satisfactory.**

2. **The uncertainties of the results of most participants have been underestimated.**