

Comparison of results of selected PTs in the food sector

Evaluation report of the 2nd comparison within CoEPT EU contract GTC1-2002-73002

Adriaan M.H. van der Veen¹, Michel van Son¹, Piotr Robouch², Ulf Örnemark³

¹Nederlands Meetinstituut, Thijsseweg 11, 2629 JA Delft, the Netherlands

²Institute for Reference Materials and Measurements, Retiesweg, B-2440, Geel, Belgium

³EQUALIS AB, Box 977, SE-751 09, Uppsala, Sweden

Delft, the Netherlands, 30 August 2005



Copyright Notice

This work is protected under copyright by the enterprises engaged in the work on this project. The work is supported in part by the European Commission under contract number GTC1-2002-73002, “Comparability of the operation and evaluation protocols of European proficiency testing schemes”.

Extracts of this report may be reproduced provided that the source is acknowledged and the extract is not taken out of context.

Summary

The 2nd comparison in the project “Comparability of the operation and evaluation protocols of European proficiency testing schemes” (CoEPT) was carried out between February and November 2004. The aim of this 2nd comparison was to see how the statistical protocols of the participating PT-providers perform under practical conditions, which are as close as possible to the normal conditions of the respective proficiency tests. The aim of the 2nd comparison is to see whether the results of the 1st comparison are confirmed when using the same samples in the proficiency testing schemes of the participating PT-providers.

A freshly prepared CRM was used as PT-material and made available to the PT-providers participating in the project. This approach was chosen because it was not possible to find an existing CRM in a sufficient number of units.

This 2nd comparison has clearly demonstrated that it is possible to make useful inferences about results obtained on the same sample measured by two different laboratories. The implied assumptions concerning metrological traceability of the results, which is a prerequisite for being able to make a useful comparison, have proved to be valid for the PTs involved in the project. That this conclusion can be extended to other PTs is likely, but subject to proof.

Generally there was good agreement between any pair of assigned values in a dataset for a given parameter. From this agreement, it can be concluded that there is generally good agreement between the PTs of the participating PT-providers. Larger differences exist in the standard deviations and uncertainties estimated from the data reported by the participating laboratories, but they simply reflect differences in (average) performance. In conclusion, there is good comparability across the participating PTs for the characterisation of milk powder.

The biggest differences between the PT-providers are observed in the establishment of the standard deviation for performance rating and the evaluation of measurement uncertainty associated with the assigned value. The assigned values from the PT-providers agree usually with the reference value established for the certified reference material (CRM) used in this comparison. This fact not only leads to the conclusion that the approaches of the PT-providers for establishing an assigned value are valid, but also that the evaluation of uncertainty is at an acceptable level.

A common protocol, based on ISO 13528 has been used to evaluate the datasets for all parameters of all PT-providers. This evaluation revealed also differences between the datasets, in particular in the level of scatter. It also made clear that the different approaches for establishing the standard deviation for performance rating come *in addition to* different levels of scatter.

Table of contents

Summary	III
Table of contents	IV
List of figures	VI
List of tables	VI
List of symbols	VI
1 Introduction	8
2 Design of the comparison	9
2.1 Samples	9
2.2 Measurands	9
2.3 Reporting requirements	9
2.4 Co-ordination and evaluation	10
2.5 Participating PT providers	10
2.6 Interpretation of results	10
3 Organisation & logistics	12
3.1 Outcome of the 2 nd CoEPT workshop	12
3.2 Ordering procedure	12
3.3 Deadlines	12
3.4 Documentation to PT providers	13
3.5 Report form for the 2 nd intercomparison	13
4 Evaluation of datasets using a common approach	15
4.1 Common protocol	15
4.2 Data evaluation of selected data sets	16
5 Results	20
5.1 Ash content	20
5.2 Fat content	22
5.3 Lactose content	24
5.4 Moisture content	26

5.5	Protein content	28
6	Discussion	30
6.1	Specification of measurands and comparability	30
6.1.1	Empirical methods	30
6.1.2	Same underlying measurement reported differently	30
6.1.3	Different measurands	31
6.1.4	Different quantities/units of measurements used for reporting of results	31
6.2	The “assigned value”	31
6.2.1	Reference values	31
6.2.2	Consensus values	32
6.3	Uncertainty of assigned values	32
6.4	The standard deviation for PT assessment	33
6.5	Initial data treatment	33
6.6	Evaluation of datasets using the common protocol	33
6.7	Comparing results across PTs	34
6.8	Results from the 5 PTs	35
7	Conclusions	36
8	References	37

List of figures

Figure 1: Robust consensus means and standard deviations for protein content	17
Figure 2: Robust consensus means and standard deviations for fat content	17
Figure 3: Robust consensus means and standard deviations for lactose content	18
Figure 4: Robust consensus means and standard deviations for moisture content	18
Figure 5: Robust consensus means and standard deviations for ash content	19
Figure 6: Assigned values for ash content	20
Figure 7: Overview of results for ash content	21
Figure 8: Assigned values for fat content	22
Figure 9: Overview of results for fat content	23
Figure 10: Assigned values for lactose content	24
Figure 11: Overview of results for lactose content	25
Figure 12: Assigned values for moisture content	26
Figure 13: Overview of results for dry matter	27
Figure 14: Assigned values for protein content	28
Figure 15: Overview of results for protein content	29

List of tables

Table 1: List of PT providers	10
Table 2: Schedule for 2 nd intercomparison	12
Table 3: Robust means and standard deviations in accordance to ISO 13528, algorithm A	16
Table 4: Consensus values, assigned values, and standard deviations for ash content (wt-%)	20
Table 5: Consensus values, assigned values, and standard deviations for fat content (wt-%)	22
Table 6: Consensus values, assigned values, and standard deviations for lactose content (wt-%)	24
Table 7: Consensus values, assigned values, and standard deviations for moisture content (wt-%)	26
Table 8: Consensus values, assigned values, and standard deviations for protein content (wt-%)	28
Table 9: Methods used for parameters in milk powder	31

List of symbols

<i>CI</i>	Confidence interval
<i>p</i>	number of laboratories

RSD_R	relative standard deviation (of an interlaboratory study) under reproducibility conditions
s	standard deviation
s^*	robust standard deviation (ISO 13528)
s_L	between-laboratory standard deviation (ISO 5725-2:1994)
$sMAD$	standard deviation of the median, estimated as $\frac{MAD}{0.6745}$
s_{PT}	standard deviation used for performance assessment
s_R	reproducibility standard deviation (ISO 5725-2:1994)
s_r	repeatability standard deviation (ISO 5725-2:1994)
u	standard uncertainty
U	expanded uncertainty
x_i	result of laboratory i
x_{cv}	consensus value
x_{med}	median
\bar{x}	mean value
x_{PT}	assigned value
x_{ref}	reference value
Z	Z -score

1 Introduction

The 2nd intercomparison in the project “Comparability of the operation and evaluation protocols of European proficiency testing schemes” (CoEPT) was carried out between February and November 2004. The objectives of the intercomparison were:

- to investigate the statistical protocols of the proficiency testing (PT) providers under conditions, which are as close as possible to the normal conditions of the respective PT scheme;
- to investigate whether the results of the 1st intercomparison are confirmed when using one or more common test materials in the PT schemes of the participating providers.
- To evaluate the equivalence of reference and consensus values as normally obtained in the PTs
- To compare the performance and rating of the participating laboratories

A PT scheme is run on the basis of a (documented) design, hereafter referred to as “statistical protocol”. This statistical protocol stipulates what data are gathered, for what purpose, and how they will be evaluated, and what type of conclusions can be drawn from the PT. The typical contents of such a protocol are given in ISO Guide 43-1 [1]. The basis for this intercomparison, as well as for the protocol [2] is given in [3].

In the 1st intercomparison, emphasis was put on the comparability and robustness of the statistical protocols used by the participating PT providers. A selection of datasets from different PT schemes was provided for this purpose. One of the weaknesses of the approach of the 1st intercomparison is that relatively little context information concerning the data is available. As mentioned in the report from the 1st intercomparison [3], this lack of context information sometimes impaired the evaluation of the data in accordance with the statistical protocol of the PT provider. This issue is not present in this 2nd intercomparison, as the PT provider was provided with samples that were to be distributed to the participants in the PT, just as any other sample. In most cases, PT providers opted for distributing the samples from this project as additional sample, rather than using it instead of one of the regular samples.

This report covers the results of the 2^d intercomparison for the Food Sector, in which five PT providers participated. One common test material, a roller dried whole milk powder, has been used. The food sector is of course much larger than only this matrix, but in the frame of this project, it is impossible to cover even what may be called “a representative subset” of the relevant matrices or measurands therein.

There are many legislative aspects involved in the food sector and a European Commission regulation exists for the analysis of milk powder. It is however, beyond the scope of this project to discuss those issues.

The idea behind the project CoEPT is to demonstrate how comparability across proficiency tests can be assessed, taking four important sectors as example. The other three sectors are drinking water, soil, and occupational hygiene.

The draft protocol for this intercomparison was discussed during the 2nd Workshop in Geel (October 2003) and provided as draft in June 2004. The final version was made available in December 2004 [2]. The changes between the draft and final protocol were solely editorial.

2 Design of the comparison

2.1 Samples

The test material was a roller dried whole milk powder, manufactured and provided by MUVA Kempten (Germany) under the identity “MUVA RM 56”. It was supplied in aluminium bags that had been sealed under inert gas atmosphere. Each bag contained roughly 50 g.

Information concerning the manufacturing process, homogeneity, and the property values assigned to the test material by the manufacturer, can be found in the certificate reproduced in Annex 1.

2.2 Measurands

The quantities subject to measurement are mass fractions of lactose monohydrate, fat, dry matter, protein and ash. To improve comparability of results from the five PT providers, the following specifications apply to [this report](#):

- The names used for the measurands are those found on the test material certificate;
- Results are reported in the unit g/100 g, i.e. in per cent;
- Results for Nitrogen is converted to Protein by multiplying with a recommended factor of 6.38;
- Results for Lactose are recalculated to lactose monohydrate **according to [insert equation]**;
- Results for Moisture are converted to Dry matter **according to [insert equation]**. The dry matter should not include the crystalline water of lactose monohydrate.

2.3 Reporting requirements

The following information was requested from the participating PT provider:

- (1) Results from data evaluation as normally undertaken by the PT provider;
- (2) Assigned value (consensus and reference value), and its associated uncertainty;
- (3) Results from outlier detection (if applicable); indication of ground for rejection;
- (4) Performance rating of the laboratories; if such a rating is not foreseen please indicate on the report forms how the laboratories are instructed to interpret the results.

The evaluation of the data, as described in this report will include, amongst other things, the following:

- (1) A review of the assigned values, in view of their uncertainties; this review includes a comparison among the proficiency tests, where appropriate complemented by a comparison with estimates obtained by other statistical techniques;
- (2) Differences between reference and consensus values, if the latter are used when a reference value is considered to be invalid or otherwise unusable for the purpose;
- (3) Direct comparison of the performance ratings; in particular laboratories that seem to have performed poorly will be looked at;
- (4) Comparison of performance requirements; if there are differences, these will be considered as facts rather than as subject to debate; the underlying deliberations are not at debate in CoEPT, but for the project objectives, these differences are relevant;
- (5) Comparison of data screening methods and results;

- (6) Compilation of stated uncertainties of assigned values, and a comparison with values obtained from some commonly applied techniques for deriving these uncertainties from comparison data; no assessment on the impact of these differences on performance rating, as with the current rating schemes, there is no impact.

2.4 Co-ordination and evaluation

The food sector was coordinated by EQUALIS. The evaluation of the 2nd intercomparison was done jointly by NMi-VSL, IRMM and EQUALIS.

2.5 Participating PT providers

The following five PT providers were project partners in the 2nd intercomparison (table 1).

Table 1: List of PT providers

Acronym	Contact person	Address	E-mail
CECALAIT	Mr Philippe Trossat	BP 129, FR-39802 Poligny, France	ph.trossat@cecalait.fr
RELACRE	Ms Ana Sofia Marques Ms Ana Maria Duarte Ms Raquel Candeias	Rua Filipe Folque, 2-6 Dt., PT-1050-113 Lisbon, Portugal	ctr.eci@relacre.pt
QM	Ms Tracey Noblett Mr Wayne Gaunt	Quality Management, Minerva House, 63 Hornby Street, Bury BL9 5BW, United Kingdom	tracey@qualitymanagement.co.uk wayne@qualitymanagement.co.uk
FAPAS	Ms Kate Wilkinson	Central Science Laboratory, Sand Hutton, York YO41 1LZ, United Kingdom	k.wilkinson@csl.gov.uk
MUVA	Mr Ulrich Leist Mr Michael Carl	MUVA Kempten, Postfach 2025, DE-87435 Kempten, Germany	ulrich.leist@muva.de michael.carl@muva.de

The link between the PT providers and their respective results and comments is not disclosed in this report. Instead, the following codes are used: F01, F02, F03, F04 and F05.

2.6 Interpretation of results

The first parameter to be compared is the assigned value. For each parameter, there is a reference value and an associated uncertainty available. The difference between the assigned value of a particular PT and the reference value can be regarded as a measure for the comparability of the PT with this independent reference. The difference is to be compared with its associated uncertainty, and as a criterion the following condition can be formulated

$$|x_{PT} - x_{ref}| \leq k \sqrt{u^2(x_{PT}) + u^2(x_{ref})} \quad (1)$$

or, equivalently,

$$\frac{|x_{PT} - x_{ref}|}{k \sqrt{u^2(x_{PT}) + u^2(x_{ref})}} \leq 1 \quad (2)$$

This ratio is commonly known as E_n . When $E_n > 1$, there can be two possible reasons

1. the difference $x_{PT} - x_{ref}$ is too large
2. the uncertainty $u(x_{PT})$ is too small

or of course a combination thereof. The presumption is made that the uncertainty associated with the reference value is credible.

The uncertainty associated with x_{PT} depends, when the assigned value is a consensus value, on the number of participants, among others. Generally speaking, the larger the number of participants, the smaller the uncertainty $u(x_{PT})$ becomes. For the interpretation of the difference $x_{PT} - x_{ref}$, this influence is not an issue – it can be expected that the agreement between the reference value and the assigned value becomes better when the number of participants increases. The assumption is then made that all measurement results are metrological traceable, which is of course subject to verification.

With regard to the agreement of the uncertainty u_{PT} , it must first be recalled that the uncertainty associated with an uncertainty is typically between 10 – 30% relative. Furthermore, differences may come from

1. differences in the number of participants
2. differences in the (average) performance of the participants
3. evaluation method

The evaluation methods have been compared in the 1st comparison [3].

3 Organisation & logistics

3.1 Outcome of the 2nd CoEPT workshop

The results of the 1st intercomparison [3] were discussed with the PT providers during the 2^d workshop (Geel, 26-28 October 2003). On this occasion, the food sector agreed on the choice of test material and on the fundamentals of the logistics for the 2nd intercomparison [4]. To summarise:

- ◆ Investigations by EQUALIS and LGC had shown that no suitable milk powder RM/CRM was available from large manufacturers [5]. The PT providers had no additional information in this respect. Two offers - from within the WG – for a test material were considered. Following detailed discussions, the WG concluded that the type of materials produced regularly by MUVA would meet the requirements for homogeneity, stability, cost, composition, and reference values. Subsequently the WG agreed to use the MUVA RM 56 material (see below).
- ◆ The 2nd intercomparison should be executed in the frame of the providers' normal PT activities. Participants should be informed about the use of their results in CoEPT after that results were collected but before processing them in CoEPT. This approach was later agreed to by all WGs and a common information leaflet was provided [6].
- ◆ Due to the low cost for the test material, the CoEPT consortium granted that all estimated participants from the food sector, i.e. about 220, could be included. Since FAPAS and QM provide larger sample sizes in their schemes than does MUVA, a total of 300-350 samples would be needed.
- ◆ MUVA was to send the samples unlabelled to the other providers, whom would then perform labelling and distribution according to their normal procedures. MUVA also agreed to send samples of the test material so that each provider could investigate it beforehand in accordance with in-house procedures. Sample dispatch could start 1st February 2004 and the PT exercises must be completed by 30 September that year.
- ◆ The WG concluded that a similar format of the documents as in the 1st intercomparison be suitable. The evaluation must take into account that reference values may be based on specific methods, which should be listed in the protocol. The protocol must enable replicate information to be collected (up to three replicates used). The provider will include all raw data in their reports to EQUALIS.

3.2 Ordering procedure

Beginning of February 2004, each provider could order the necessary number of samples from MUVA according to an agreed procedure [7].

3.3 Deadlines

The schedule for this intercomparison followed the schedule of the PT providers. The major deadlines as indicated in the protocol [2] were (see table 2).

Table 2: Schedule for 2nd intercomparison

Deadline	Deliverable	Responsible	To be sent to
30.09.2004	PT rounds completed	PT providers	-
29.10.2004	Evaluation of data, report forms, PT report	PT providers	SCs
21.12.2004	Draft reports (for each sector one)	NMi VSL	PT providers, SCs
10.02.2005	Comments on draft reports	PT providers, SCs,	NMi VSL

Deadline	Deliverable	Responsible	To be sent to
		NCs	
01.03.2005	Final draft reports	NMi VSL	PT providers, SCs, NCs

The food PT providers received information regarding ordering of the test material [7] and the protocol [2] beginning of February 2004. An Excel report form [8] and accompanying information were distributed by email 30 August. PT rounds were executed between February and October. Two providers submitted their evaluation reports prior to deadline 29 October [9,10] and the remaining three reports were received by EQUALIS between 24 November and 17 December [11,12,13]. EQUALIS forwarded the PT providers' data to NMi-VSL/IRMM, which returned a draft evaluation 4 January 2005 together with guidance for report layout.

EQUALIS prepared a first draft of this report and sent it (5 January) to the PT providers LGC and NMi-VSL/IRMM. The preliminary results and the content of the draft were discussed in connection with the 3rd CoEPT workshop and meetings (Ede, The Netherlands, 7-9 Feb. 2005).

3.4 Documentation to PT providers

The common set of documents that EQUALIS sent each PT provider consisted of:

- ◆ Protocol for the 2nd intercomparison (Word format) [2]
- ◆ Report form for the 2nd intercomparison (Excel format) [8]
- ◆ Homogeneity data for the MUVA RM 56 test material (Excel format) [14]

The report form is further commented on below. The properties of the test material are described in section 2.1 and in Annex 1.

3.5 Report form for the 2nd intercomparison

The Excel report form served as basis for the evaluation and forthcoming report for the food WG. It could also be considered as a checklist to see that the information transferred from the PT providers' own databases complied with that requested. The report form contained seven sheets: "PT provider info", "Summary", and one sheet for each analyte.

In the sheet "PT provider info", providers could indicate if there were changes to the evaluation procedure (e.g. statistical protocol) compared to the information provided in the report for the first intercomparison. Providers F02 and F05 had made no changes. Providers F01 and F03 both chose to use assigned values and assigned standard deviations derived by MUVA. The reason was the small number of participants' results. Provider F04 had changed its protocol. The consensus value was now calculated according to algorithm A in ISO 13528 [15] after elimination of outliers using Grubbs test at 5 %.

The five sheets for the measurands were prepared with columns for:

- ◆ participants i.d. (1-n) and/or lab code
- ◆ columns for up to 3 replicate measurement results
- ◆ performance statistics, e.g. z-score
- ◆ performance rating, e.g. "Satisfactory"

The PT providers were instructed to transfer any relevant information (tables, graphs etc) from their local PT database into the respective parameter sheet. Specific observations and comments about results could also be mentioned here. Any detected outliers should be clearly indicated somewhere in the information.

Finally, the sheet “Summary” was to be completed with information about the assigned values and their uncertainties, and the assigned standard deviations (see section 2.3).

4 Evaluation of datasets using a common approach

4.1 Common protocol

The discussions during the 3^d CoEPT workshop held in Ede, the Netherlands, on 8 February 2005 indicated that there was a need for evaluating the data with a common statistical protocol, to assess broadly the equivalence between the PTs participating in the project.

The protocol selected was based on ISO 5725-5:1998 [29], algorithm A, which is also found in ISO 13528 [15]. The consensus value and the robust standard deviations have been calculated using the following algorithm:

Denote the p items of data, sorted into increasing order, by:

$$x_1, x_2, \dots, x_i, \dots, x_p$$

Denote the robust average and robust standard deviation of these data by x^* and s^* . Calculate initial values for x^* and s^* as:

$$x^* = \text{median of } x_i \quad (3)$$

$$s^* = 1.483 \times \text{median of } |x_i - x^*| \quad (4)$$

Update the values of x^* and s^* as follows. Calculate:

$$\mathbf{j} = 1.5s^* \quad (5)$$

For each x_i calculate

$$x_i^* = \begin{cases} x^* - \mathbf{j} & x_i < x^* - \mathbf{j} \\ x^* + \mathbf{j} & x_i < x^* + \mathbf{j} \\ x_i & \text{otherwise} \end{cases} \quad (6)$$

Calculate the new values of x^* and s^* from:

$$x^* = \frac{1}{p} \sum_{i=1}^p x_i^* \quad (7)$$

and

$$s^* = \frac{1}{p-1} \sqrt{\sum_{i=1}^p (x_i^* - x^*)^2} \quad (8)$$

The robust estimates x^* and s^* may be derived by an iterative calculation, i.e. by updating the values of x^* and s^* several times, until the process converges. Convergence may be assumed when there is no change from one iteration to the next in the third significant figure of the robust standard deviation and of the equivalent figure in the robust average.

Furthermore, it was assumed Z -scores be used, with the following interpretation:

$|Z| \leq 2$: satisfactory performance (in graphs: green area)

$2 < |Z| < 3$: questionable performance (in graphs: yellow area)

$|Z| \geq 3$: unsatisfactory performance (in graphs: red area)

An alternative interpretation of the colours in the graphs is as follows:

Green = 95% CI¹

Green + yellow = 99% CI

Where a t -distribution is assumed. Furthermore, it is assumed that the datasets are symmetrical, which is not always the case for the parameters discussed in the next section.

4.2 Data evaluation of selected data sets

Table 3 summarises the results of selected parameters.

Table 3: Robust means and standard deviations in accordance to ISO 13528, algorithm A

		F01	F02	F03	F04	F05
Protein	x_{cv}	26.24	26.57	26.58	26.86	26.71
	s^*	0.62	0.75	1.33	0.04	0.27
Fat	x_{cv}	30.79	25.76	26.20	26.01	25.99
	s^*	8.57	0.63	0.48	0.13	0.24
Lactose	x_{cv}		35.82		36.48	37.75
	s^*		2.37		2.10	0.55
Moisture	x_{cv}		3.80	3.60	3.82	3.89
	s^*		0.39	0.26	0.14	0.13
Ash	x_{cv}		5.66	5.70		5.65
	s^*		0.11	0.00		0.07

The datasets submitted by PT-provider F01 were very small², so that these results can deviate more from those from the other PT-providers.

The results of the evaluation of the data of all PTs using the common protocol as described in section 4.1 is shown figure 1 for protein content. There is a good degree of agreement concerning the consensus values. The robust standard deviations differ appreciably.

¹ CI = Confidence interval

² The PT-provider stated it as follows: "Originally approximately 40-50 laboratories were contacted to participate in this ring trial. Unfortunately, due to the nature of the laboratories (i.e. dairy labs rather than contract testing labs) very little interest was shown in another round of proficiency testing. Subsequently we had a total of 10 labs that were interested. However after the trial had closed it was apparent that only 4 of these labs had actually performed any testing. Consequently the results are limited and relatively meaningless. Had such a limited number of results been returned in a normal round of proficiency testing the round would have been declared void."

CoEPT -- ISO 13528 evaluation

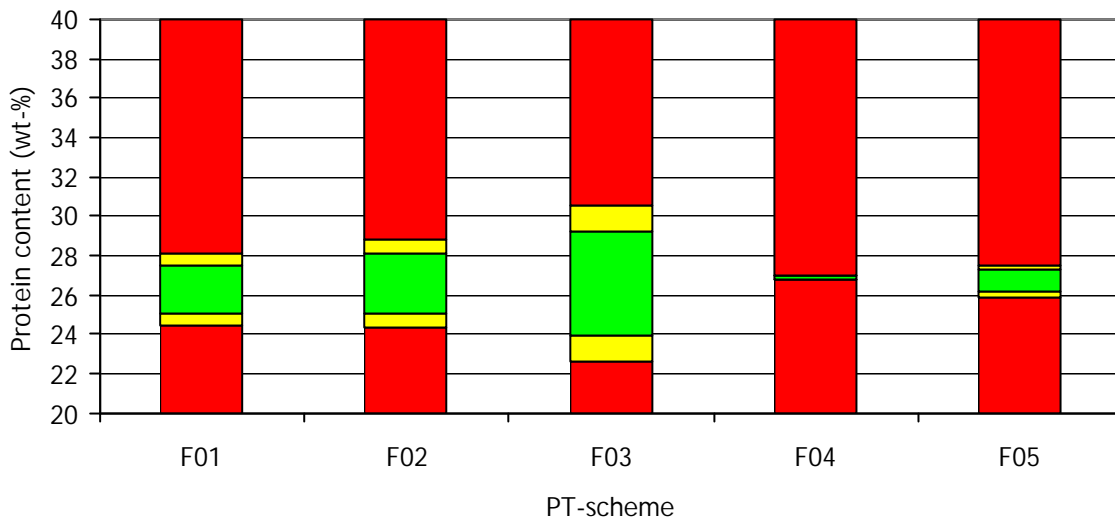


Figure 1: Robust consensus means and standard deviations for protein content

CoEPT -- ISO 13528 evaluation

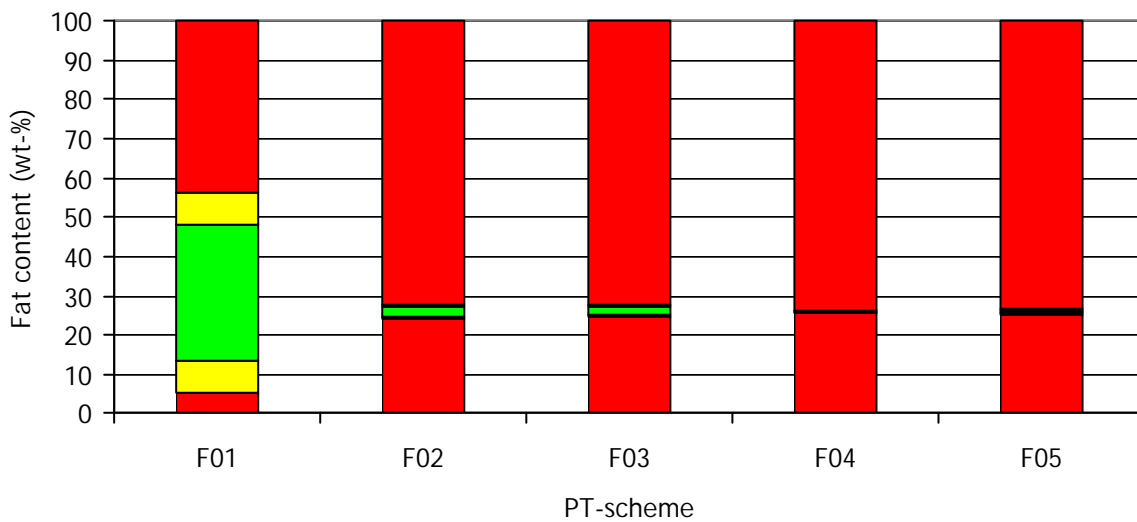


Figure 2: Robust consensus means and standard deviations for fat content

The results of the evaluation using the common protocol as described in 4.1 for fat content are shown in figure 2. The consensus values agree well. The robust standard deviation calculated from the dataset from PT-provider F01 is very large, but this is due to the fact that the dataset is very small and then even robust statistics are not capable of coming up with a realistic estimate for the standard deviation.

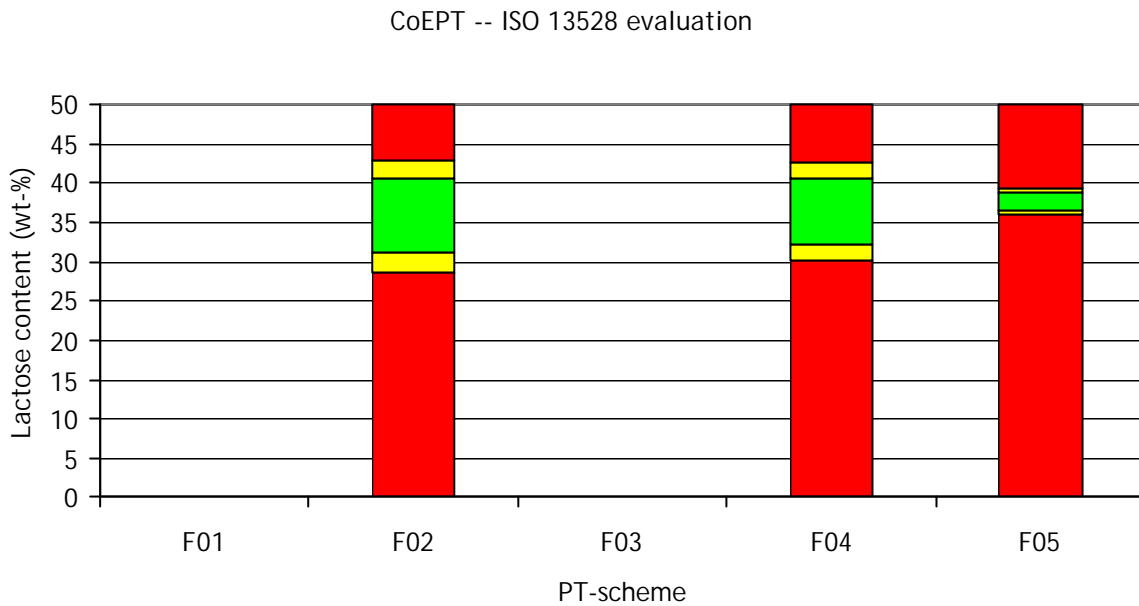


Figure 3: Robust consensus means and standard deviations for lactose content

The results of the evaluation using the common protocol for lactose content are shown in figure 3. The consensus values agree well, whereas the standard deviations calculated from the datasets show appreciable differences.

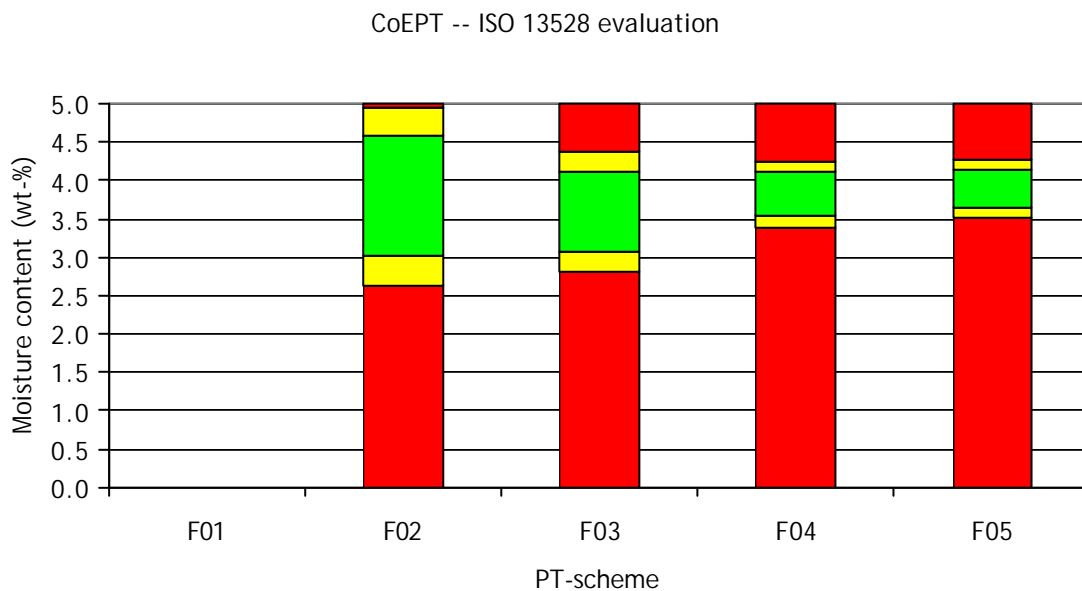


Figure 4: Robust consensus means and standard deviations for moisture content

The results of the evaluation of the datasets of moisture content are shown in figure 4. The robust standard deviations calculated from the various datasets differ appreciably, and in particular in the

datasets from PT-providers F02 and F03 the standard deviations are big in comparison to the others. The agreement among the consensus values is good.

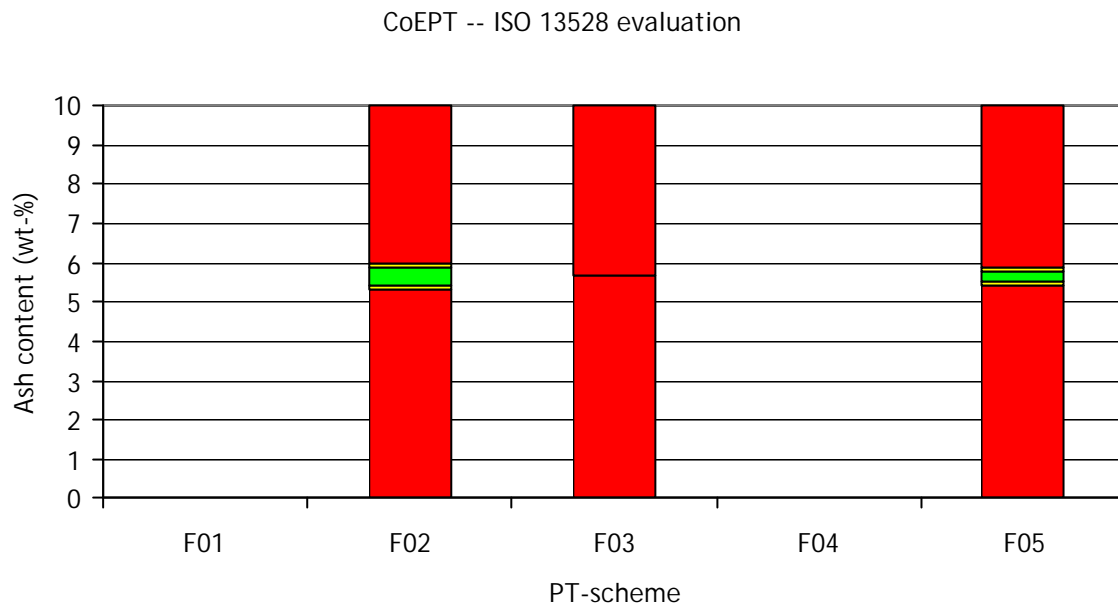


Figure 5: Robust consensus means and standard deviations for ash content

The results for ash content (figure 5) show good agreement among the consensus values. The robust standard deviations differ appreciably.

5 Results

5.1 Ash content

The assigned values and their associated uncertainties for ash are plotted in figure 6. The uncertainties are shown as expanded uncertainties with coverage factor $k = 2$, which implies approximately 95% coverage. On the left-hand side, the reference value assigned to the material (see chapter 2) is shown.

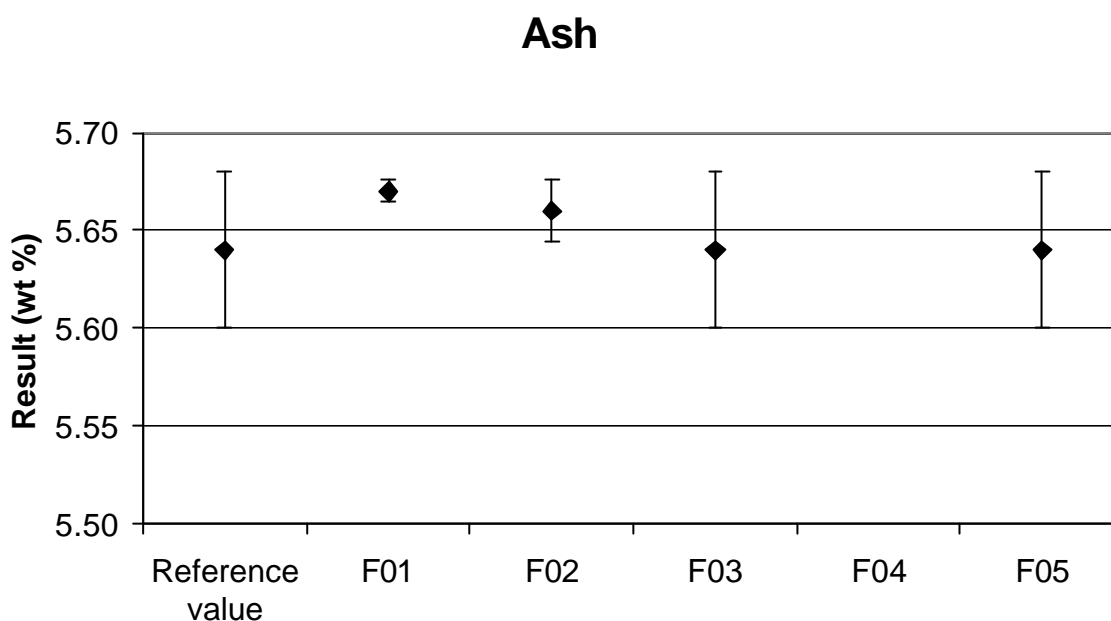


Figure 6: Assigned values for ash content

The first column in table 4 lists the PT-provider, the second and third the consensus value and its associated uncertainty, the fourth and fifth column the assigned value and its associated uncertainty, and the last two columns give the standard deviation and the target standard deviation (used for rating the laboratory performance) respectively.

The overview of the standard deviations (see table 4) reveals that there is little agreement among the standard deviations estimated from the data.

Table 4: Consensus values, assigned values, and standard deviations for ash content (wt-%)

PT	x_{cv}	$U(x_{cv})$ ($k=2$)	x_{av}	$U(x_{av})$ ($k=2$)	s	s_{PT}
F01	5.67	0.02	5.67	0.0058	0.05	
F02	5.66	0.02			0.074	
F03	5.7		5.64	0.04	0.228	
F04						
F05	5.64	0.02	5.64	0.07		

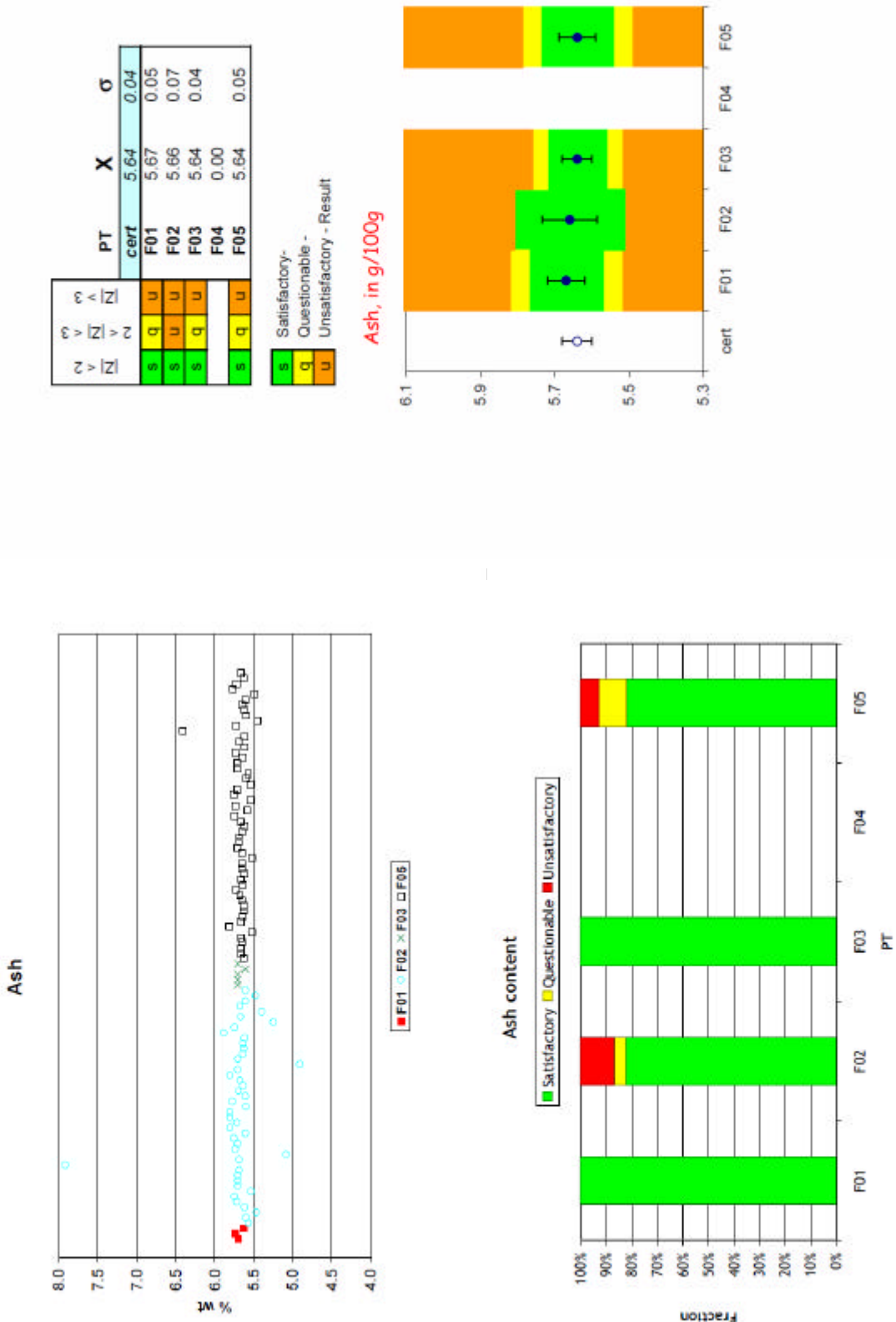


Figure 7: Overview of results for ash content

5.2 Fat content

The assigned values and their associated uncertainties for fat content are shown in figure 8. The agreement between the reference value (on the left-hand side of figure 8) and the assigned values of the PT-providers is good for most PTs, with the exception of the PT F02.

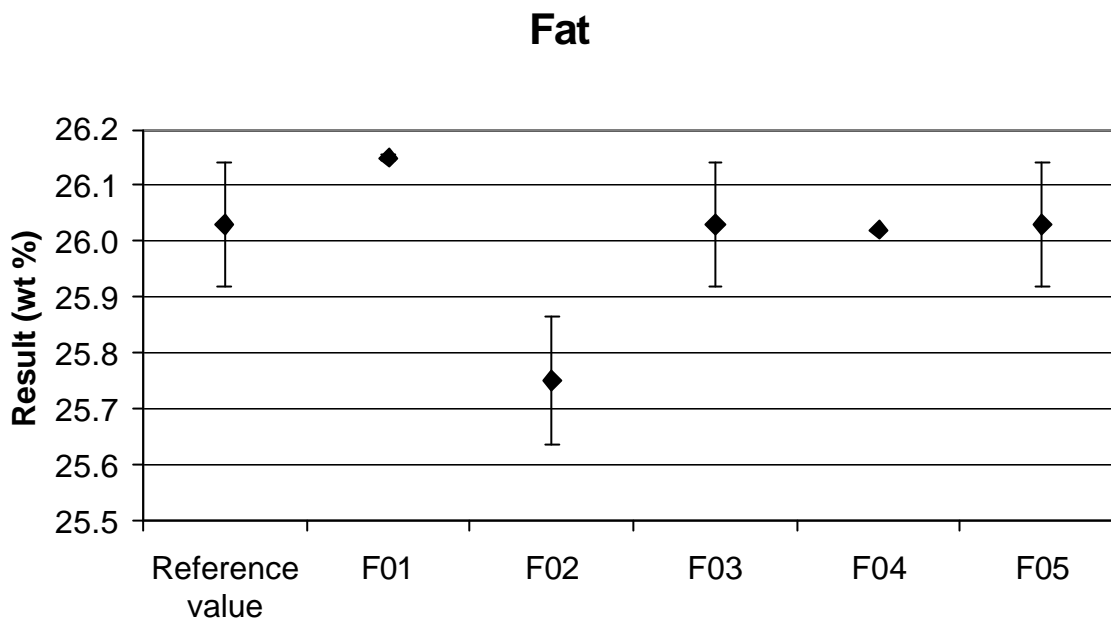


Figure 8: Assigned values for fat content

Even without information concerning the uncertainty of measurement from PT-providers F01 and F03, it can be concluded that the assigned values are consistent with the reference value of the PT-material.

The standard deviations estimated differ appreciably (see table 5), and the same applies to the estimated uncertainties.

Table 5: Consensus values, assigned values, and standard deviations for fat content (wt-%)

PT	x_{cv}	$U(x_{cv})$ ($k=2$)	x_{av}	$U(x_{av})$ ($k=2$)	s	s_{PT}
F01	25.96	0.025	26.15	0.0039	0.10	
F02	25.75	0.116	NA	NA	0.51	
F03	26.2	0	26.03	0.11	4.37	
F04	26.02	0	0	0	0.09	
F05	25.99	0.070	26.03	0.06	0	

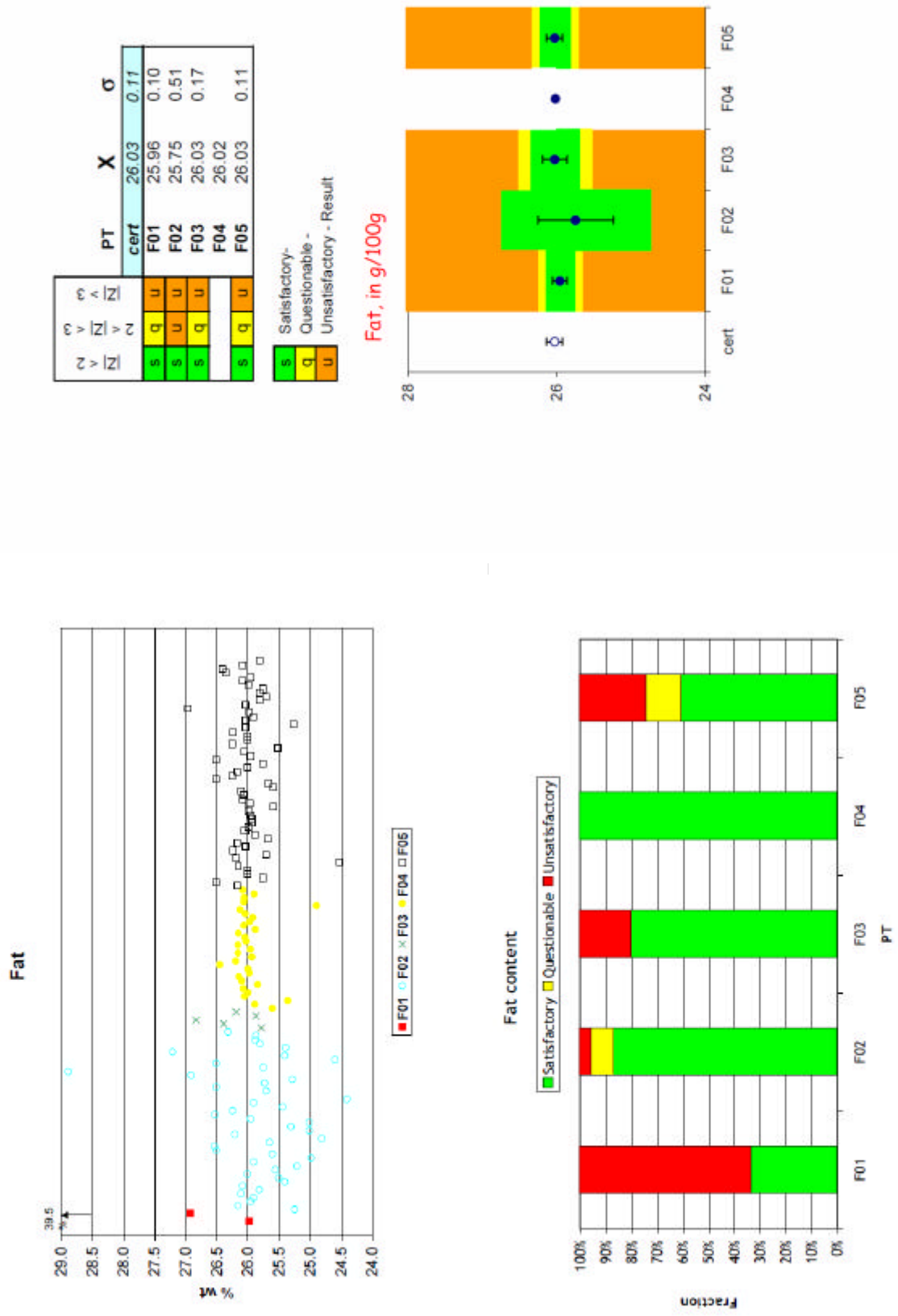


Figure 9: Overview of results for fat content

5.3 Lactose content

The assigned values and their associated uncertainties for lactose content are shown in figure 10. The agreement between the reference value (on the left-hand side of figure 10) and the assigned values of the PT-providers is good for most PTs, with the exception of the PT F02.

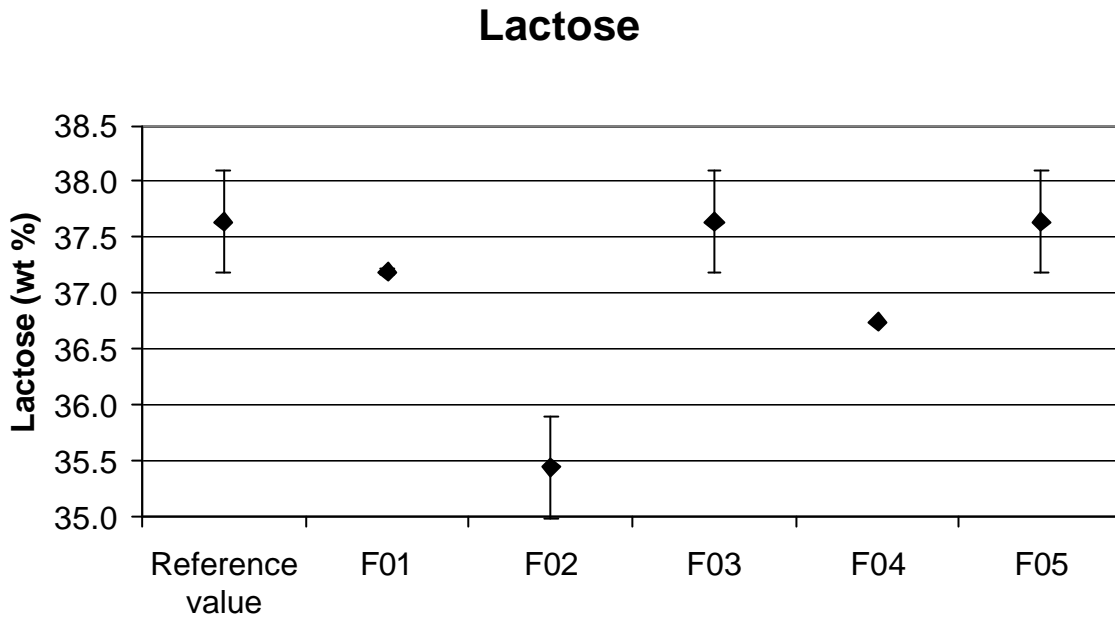


Figure 10: Assigned values for lactose content

With the exception of the assigned value of PT-provider F02, there is good agreement among the assigned values. The assigned values of PT-providers F01, F03, and F05 agree with the reference value assigned to the lactose content of the PT-material. Whether this is the case for PT-provider F04 cannot be judged, as the uncertainty data, which is lacking, is decisive.

There is little agreement among the standard deviations estimated (see table 6).

Table 6: Consensus values, assigned values, and standard deviations for lactose content (wt-%)

PT	x_{cv}	$U(x_{cv})$ ($k=2$)	x_{av}	$U(x_{av})$ ($k=2$)	s	s_{PT}
F01	23.82	2.00	37.19	0.029	0.56	
F02	35.44	0.46	NA	NA	1.134	
F03	41.14	0	37.64	0.45	18.51	
F04	36.7456	0			1.6	
F05	37.964	0.44	37.64	0.45		

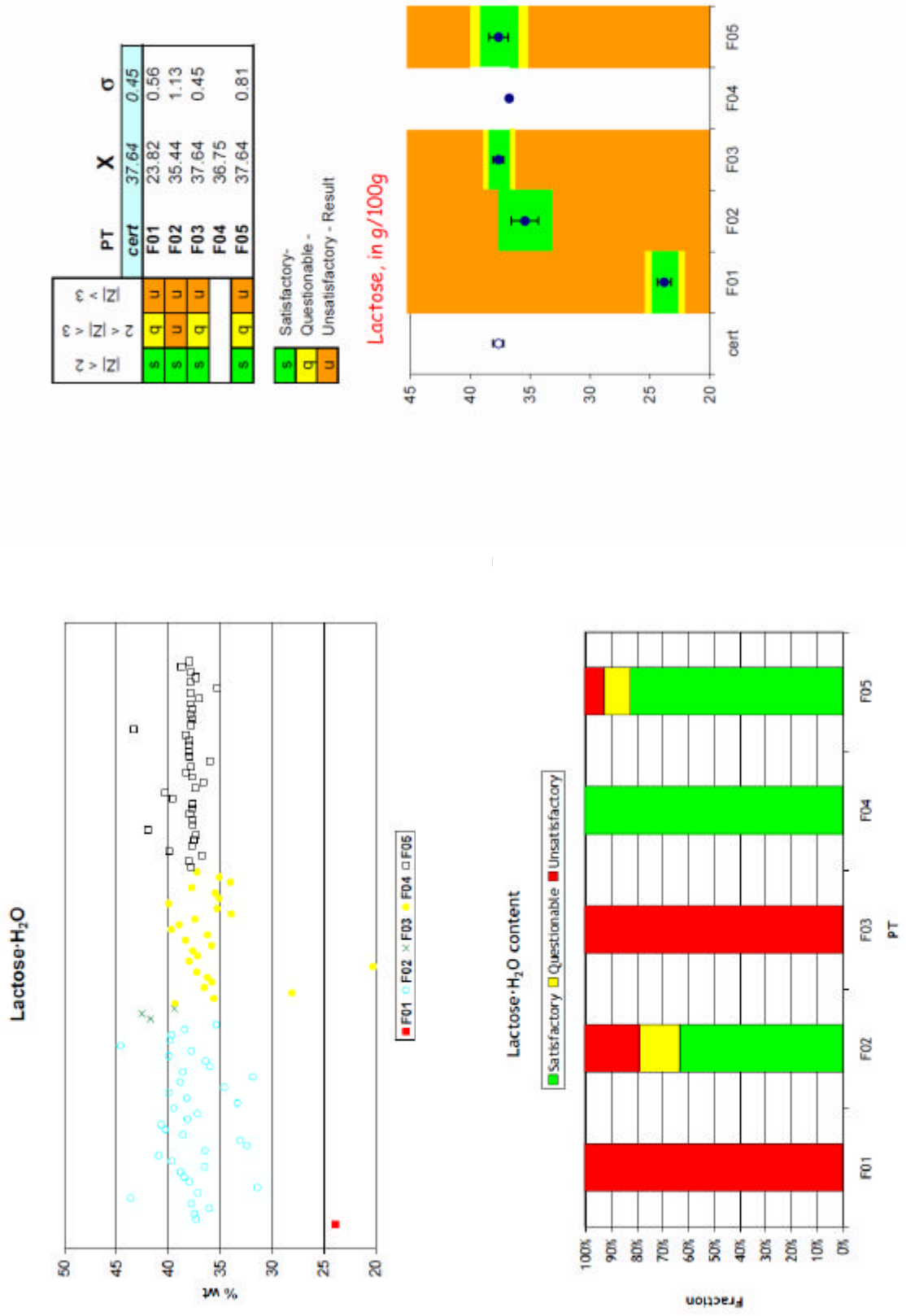


Figure 11: Overview of results for lactose content

5.4 Moisture content

The assigned values and their associated uncertainties for moisture content are shown in figure 12. The agreement between the reference value (on the left-hand side of figure 12) and the assigned values of the PT-providers is good.

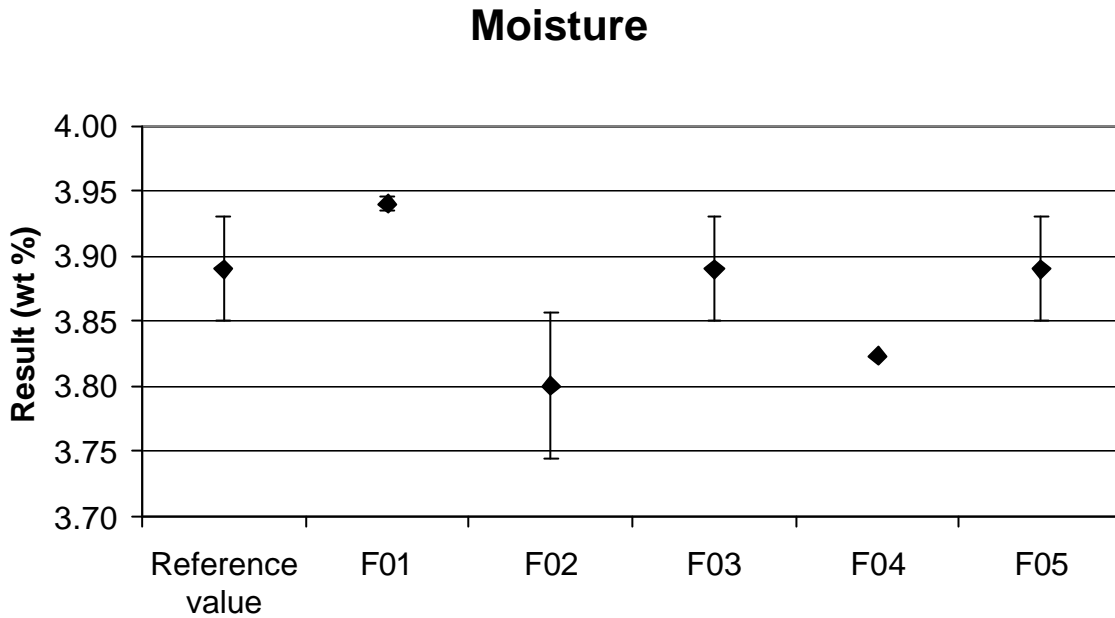
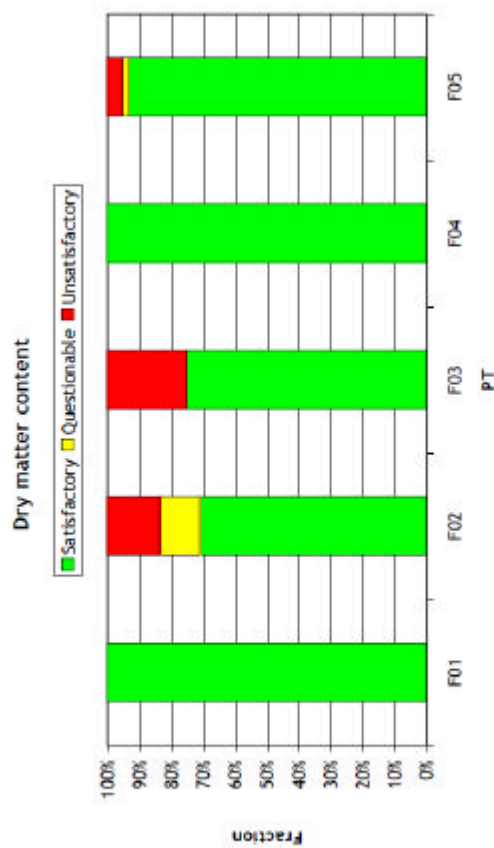
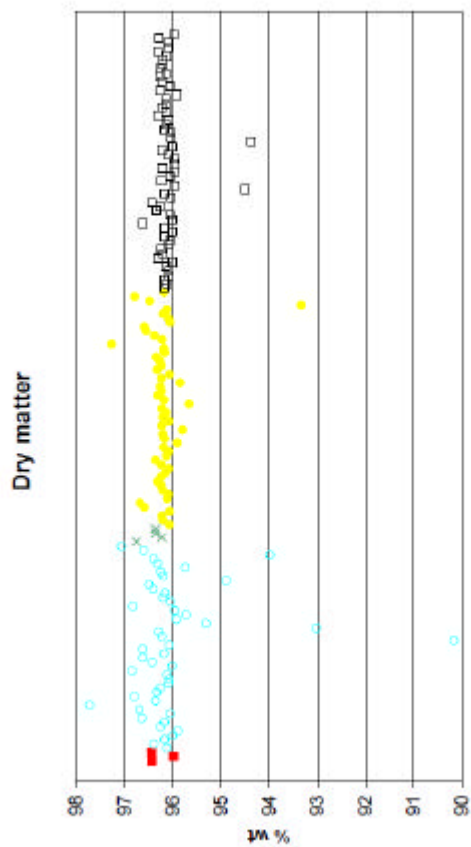


Figure 12: Assigned values for moisture content

The estimated uncertainties differ appreciably, and so do the standard deviations (table 7).

Table 7: Consensus values, assigned values, and standard deviations for moisture content (wt-%)

PT	x_{cv}	$U(x_{cv})$ ($k=2$)	x_{av}	$U(x_{av})$ ($k=2$)	s	s_{PT}
F01	3.68	0.011	3.94	0.0058	0.20	0
F02	3.80	0.056	NA	NA	0.209	0
F03	3.6	0	3.89	0.04	0.6	0
F04	3.82	0	0	0	0.1	0
F05	3.88	0.03	3.89	0.030	0	0



Z		V		N		PT		X		σ	
Z	V	Z	V	Z	V	cert	F01	F02	F03	F04	F05
S	Q	S	Q	S	Q	96.11	96.32	96.20	96.11	96.18	96.11
S	Q	S	Q	S	Q	0.11	0.20	0.21	0.17	0.14	

■ Satisfactory-
■ Questionable -
■ Unsatisfactory - Result

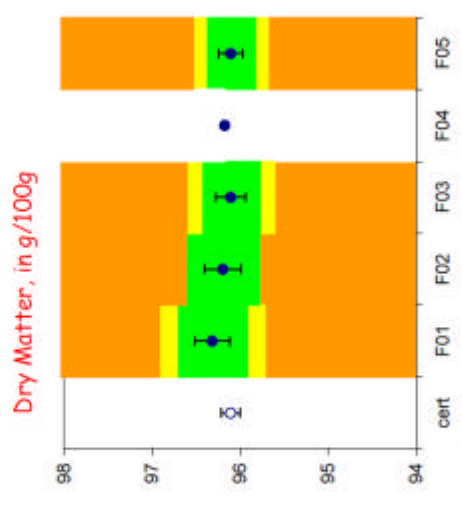


Figure 13: Overview of results for dry matter

5.5 Protein content

The assigned values and their associated uncertainties for protein content are shown in figure 14. The agreement between the reference value (on the left-hand side of figure 14) and the assigned values of the PT-providers is good for most PTs, with the exception of the PTs F01 and F03.

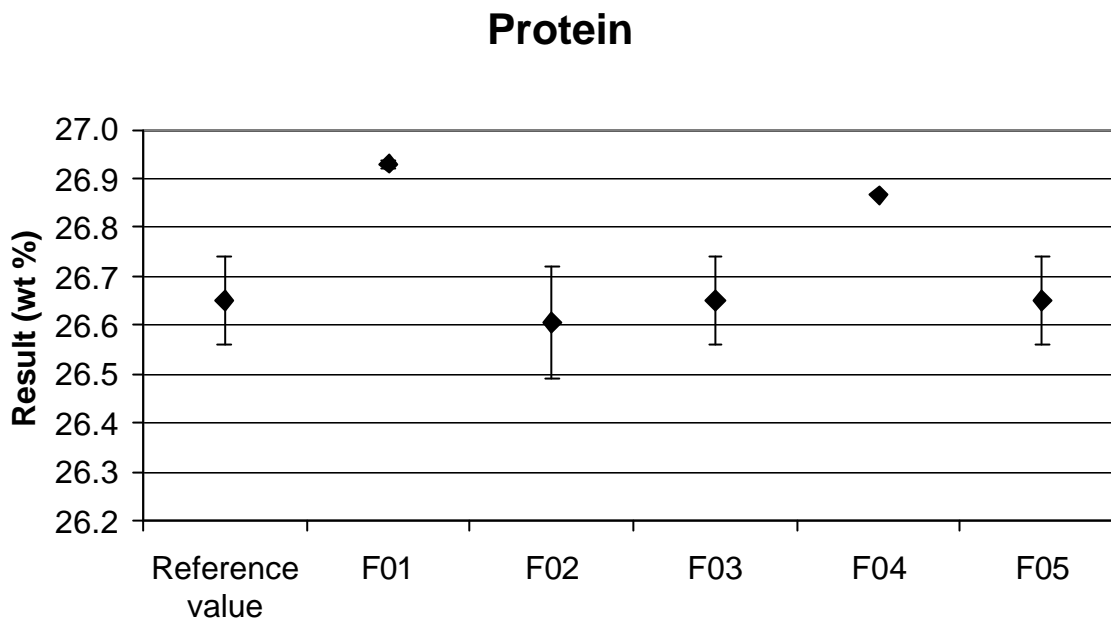


Figure 14: Assigned values for protein content

The estimated uncertainties for the consensus values agree well (see table 8), whereas the standard deviations differ appreciably.

Table 8: Consensus values, assigned values, and standard deviations for protein content (wt-%)

PT	x_{cv}	$U(x_{cv})$ ($k=2$)	x_{av}	$U(x_{av})$ ($k=2$)	s	s_{PT}
F01	26.09	0.11	26.93	0.01	0.30	0
F02	26.60	0.11	NA	NA	0.85	0
F03	26.63	0	26.65	0.09	4.44	0
F04	26.87	0	0	0	0.17	0
F05	26.66	0.08	26.65	0.08	0	0

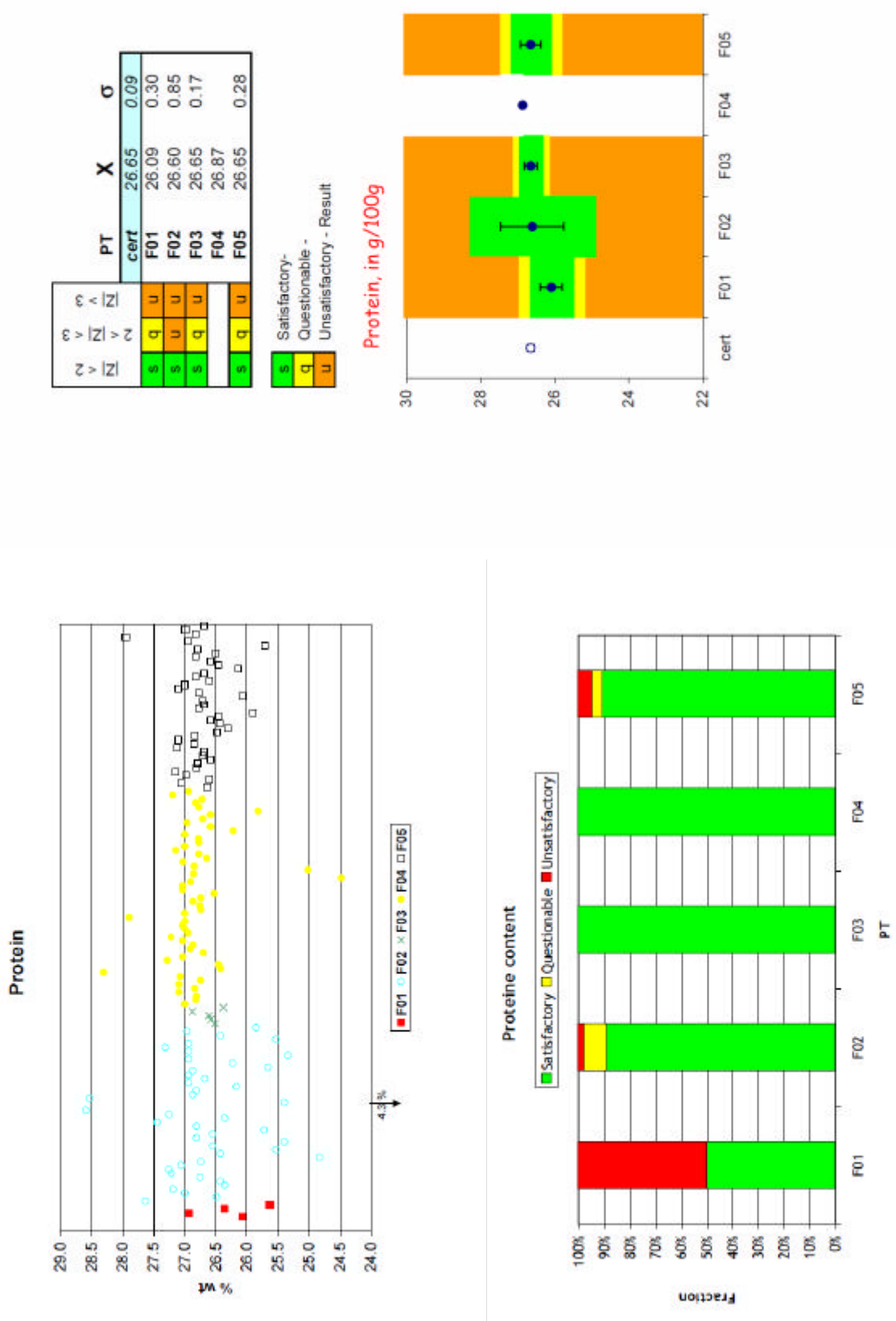


Figure 15: Overview of results for protein content

6 Discussion

6.1 Specification of measurands and comparability

For the sake of this project report, there was a need for harmonising the participants' results. The topic was discussed during the 3rd workshop [16]. The PT-providers expressed different views on how to harmonise results, suggesting using the measurand specifications

- most frequently occurring in the five PT-schemes, or
- those stated in standards, or
- those stated in the certificate from the RM producer (Annex 1).

As indicated already in section 2.2, the last option has been chosen as a good compromise.

The reasons why some of the results from the PT providers' schemes are not directly comparable are outlined in Sections 6.1.1 - 6.1.4).

6.1.1 Empirical methods

For "empirical methods", the measurand is defined by the applied procedure. Changing the procedure, or failure to comply with it (within specified limits), may lead to significantly different results. The "content of water" in biological samples can be difficult to determine simply because water is bound to the matrix in many ways. Attempts to remove it by heating may decompose other constituents resulting in errors. The specification of the measurand should, therefore, refer to the applied procedure. The complexity is illustrated in the results report [14] from MUVA where four drying temperatures between 87 and 117 °C are stated with reference to seven standards. Also for the other four measurands, specifications need to refer to a specific procedure.

6.1.2 Same underlying measurement reported differently

The most common procedure for determining organic (amine) nitrogen is that of "Kjeldahl" [17]. First, bound nitrogen is converted to ammonium ion in an acid digestion. Ammonium is then converted to ammonia, which is distilled and determined by a neutralisation titration. The classical method has been modified to enable also more highly oxidised forms of nitrogen to be determined.

So far the measurand is rather well specified. However, in food science and for nutritional purposes, it is more common to express the result as "protein content". The protein composition is roughly the same in each type of biological material. For dairy products, the protein content is obtained by multiplying the nitrogen content by 6,38, a factor, although an approximation, that has been internationally agreed upon [18,19]. A general factor of 6,25 can be used if the results are required for nutritional information (based upon the food labelling regulations). In two of the PT-providers' (F02 and F04) schemes, results are expressed as "nitrogen content" and the other three schemes as "protein content".

"Moisture content" is the new official "name" to describe the presence of water in milk powder [19, 20]. The British standard BS1743 describes both the moisture and total solids content analysis as the 'maximum loss of mass', and describes the calculations for determining these parameters. Two of the PT providers (F01 and F05) calculate the "dry matter" based on the water determination and use it in their schemes.

Also for the analyte "lactose" there are comparability problems. Providers F04 [21] and F05 refer to lactose monohydrate while results in providers' F01 and F02 schemes are said to be for anhydrous lactose [22,23]. When specifying dry matter, it is assumed by provider F05 that lactose is present as lactose monohydrate, and that the crystalline portion of water is not included in the dry matter [19].

6.1.3 Different measurands

Differences between PT providers, or a larger than anticipated variation within one scheme, can be due to the fact that participants do not measure or report values for the same measurand. Lactose (molar mass 342,30) should be differed from lactose·H₂O (molar mass (360,31)). A first compilation of lactose results showed a slightly lower average value for Provider F02 compared to the reference value. The difference (~6%) could, too a large extent, be explained by reporting results for different measurands. Provider F02 had suspected differences between participants in the scheme and therefore asked participants prior to the comparison to state their reporting format.

Provider F04 has stressed that, in principle, results for protein based on the Dumas procedure should be separated from those obtained by the Kjeldahl procedure. The former is a combustion method and gives a result for total nitrogen. However, standards and/or local protocols may already have taken such differences into account.

6.1.4 Different quantities/units of measurements used for reporting of results

Lack of harmonization can explain that measurement results are sometimes linked to different quantities, e.g. mass fraction or mass concentration. Rules for expressing results, e.g. as outlined in international standards [24] are not always adhered why values and units for the same quantity vary. In this intercomparison there seems to be no problem with respect to either quantity or unit. All quantities are claimed to be “mass fractions” with results expressed in g/100 g, i.e. in per cent.

6.2 The “assigned value”

In PT, a comparison is made between a laboratory’s result and an *assigned value*, which can be obtained by processing the data from all participants (*consensus value*). Alternatively selected results from either the participants or from an independent source can be used (*reference value*).

6.2.1 Reference values

International standards for the milk sector are now jointly published by ISO and the international dairy federation (IDF). Provider F05, who is also the manufacturer of the test material, establishes the reference values from participants that use officially recognised measurement procedures, e.g. those of ISO and IDF (table 9). Such procedures are well documented and considered “reliable”, i.e. they have higher metrological qualities (with respect to traceability and uncertainty) than those used in routine analysis.

Table 9: Methods used for parameters in milk powder

Analyte	ISO Standard	Corresponding sector standard or national reference	Remark
Fat	ISO 1736 [25]	<ul style="list-style-type: none">IDF 9C:1987	Röse Gottlieb reference measurement procedure
Protein	ISO 8968-1,2 [17]	<ul style="list-style-type: none">IDF 20-1,2 via procedure according to Vienna agreementGerman official methods § 35 L 01.00-10-1,2	Kjeldahl procedure
Ash	-	<ul style="list-style-type: none">DIN 10477 [26]German official method § 35 L 01.00-77	Drying at 525±25 °C for 2h

Lactose-H ₂ O	ISO 5765-2 [27] ¹	<ul style="list-style-type: none"> • IDF 79-2:2002 • German official method § 35 L 01.00-17 • DIN 10344 	Enzymatic method
Dry matter	ISO 5537 [20]	<ul style="list-style-type: none"> • IDF 26a:1993² 	Drying at 102 °C

In a direct comparison between proficiency tests, reference values form to some extent a problem, because it is very hard to simulate reality in this respect. Most PT providers have in their statistical protocol an “escape hatch” for the circumstance that a reference value is deemed “non-credible”. There can be various reasons for such a belief, for example because – against all odds – the reference value differs appreciably (and often significantly) from the consensus value, even if the latter is not formally established.

6.2.2 Consensus values

Pt providers F01-F04.....

6.3 Uncertainty of assigned values

The use of uncertainty in PT is still a bit underdeveloped. Although there are some clear advantages of using uncertainty in PT, the well-known *Z*-score is, for various reasons, still used. Uncertainty information usually plays a secondary role (if at all) in the performance rating. The advantages of using uncertainty information include:

- ◆ A consensus value can be weighted on the basis of the accuracy of the laboratory result,
- ◆ Performance rating becomes individual, instead of one criterion for all,
- ◆ PT becomes a tool in getting the evaluation of measurement uncertainty “right”.

Many PTs are still operated without this information, for reasons like:

- ◆ Lack of interest among the participants
- ◆ Lack of credible uncertainty data
- ◆ Sticking to well-established “old” habits

Each of these arguments has some validity, and influences what a PT provider can do.

In this intercomparison, the uncertainty associated with the assigned value plays a minor role. It has been asked for because, e.g., on request of a participant, this information needs to be available (see, ISO Guide 43-1 [1]). For assessing performance across proficiency tests, the uncertainties of the assigned values are needed as well. The estimated uncertainties are compared with one another, but certainly in the context of this intercomparison it is impossible to give a “definitive” evaluation of the uncertainty of an assigned value.

A European regulation has to be implemented not later than 1st Jan. 2007. That directive will require food laboratories to be accredited and, hence more attention will be paid to uncertainty on reported results [28].

¹ The standard procedures for determination of lactose are described in ISO 5765 [Error! Bookmark not defined.]. They are not very precise (repeatability $r=3\%$ and reproducibility $R=6\%$) why working groups are currently developing new methods based on HPLC and on differential ph.

² An upgraded version (IDF 26b) exists.

6.4 The standard deviation for PT assessment

The between-laboratory standard deviation is a parameter often calculated in PT. In some cases, it is also used for performance rating (the z score), whereas in other cases a fixed standard deviation is used for the latter purpose. Such a standard deviation can be based on customer, or regulatory, or other requirements. It can also be based on an agreement between the participants and the PT provider about what is deemed to be “satisfactory performance”.

To be added.....

6.5 Initial data treatment

Before calculating an assigned value and the standard deviation for PT assessment, the provider normally investigates the submitted results, e.g. to identify, e.g.:

- ◆ Outliers/stragglers
- ◆ Bi- or multimodality in the probability density function of the data
- ◆ Values below the reporting limit

In practice, laboratories have a policy for reporting quantitative results. The policy often includes a “reporting limit” (RL), which means that if a result is below the RL, it is indicated as “< RL”. This type of reporting can be acceptable for a laboratory’s customer, but in PT it creates additional problems for the provider. When a PT provider simply ignores these values, it may happen that the consensus value gets biased. In fact, the resulting consensus value will be too high, as the lower end of the probability density function is “cut off” at the RL. In practice, RLs may differ from laboratory to laboratory, which makes the evaluation of data even more complex.

Add information from PT provider’s result reports.....

6.6 Evaluation of datasets using the common protocol

The use of a common statistical protocol as described in 4.1 as an aid to interpret the datasets from the 10 PT-providers has proved to be very useful. Although in the first comparison it was concluded that there existed no big differences in the statistical protocols of the participating PT-providers, in particular the interpretation of the various standard deviations calculated from data and used has proved to be sometimes difficult [3]. The statistical protocol chosen has been based on algorithm A of ISO 5725-5 [29] and can also be found in the new standards ISO 13528 [15]. Furthermore, it has been assumed that Z -scores be used, which for this sector is usually the preference of the PT-providers too. The common protocol should however not be misinterpreted as being better, preferred, or otherwise superior to the statistical protocols of the PT-providers.

The validity of the approach of using a common protocol lies in the following observations:

1. Calculating a robust consensus value and a robust standard deviation is possible from any dataset in this sector
2. All consensus values obtained can be compared, if necessary taking into consideration their respective uncertainties (which can be calculated as described in, e.g., ISO 13528 [15])
3. All robust standard deviations can be compared
4. There is no 'noise' in the interpretation of consensus values or standard deviations.
5. The approach is relatively insensitive to reporting errors, straight blunders, and outliers.

6.7 Comparing results across PTs

The evaluations of the datasets of all PT-providers for all parameters revealed that, with the exceptions recorded in section 4.2, generally the consensus values agree, and therefore the results of the ('satisfactory' performing) PT-participants. The degree of agreement between two laboratories in a single PT is characterised by the reproducibility as defined in ISO 5725-2 [30],

$$R = 2\sqrt{2}s_R \quad (9)$$

where in this instance for the reproducibility standard deviation s_R the robust standard deviation of the common protocol should be substituted. The reproducibility can be interpreted as the maximum deviation between any two results of two PT-participants that is still (statistically) insignificant at a given level of confidence, in the case of ISO 5725, 95%. If two PTs have (significantly) different levels of scatter, characterised in the case of the common protocol by two different values for the robust standard deviation, the expression for the maximum difference across the proficiency tests becomes

$$R_{12} = 2\sqrt{(s_1^*)^2 + (s_2^*)^2} \quad (10)$$

It can easily be verified that if both robust standard deviations are equal, then the expression for the reproducibility as taken from ISO 5725-2 is obtained. Equation (10) is only valid when the consensus values are equal. Furthermore, it is assumed that the metrological traceability of the results of the PT-provider meets elementary requirements, such as those stipulated in clause 5.6 of ISO/IEC 17025 [31]. This 2nd comparison in the CoEPT project is unique in its approach that in different PT-schemes, the same samples have been used, so that the condition concerning the consensus values is reasonably met. The assumption concerning the metrological traceability is harder to verify, but if the PT-participants regularly participate in PTs in which the provider gives attention to this aspect in the design of its protocol, it is reasonable to assume that this assumption is also valid.

The agreement among the consensus values is exactly for this reason so important, that it gives evidence that something useful can be said about the comparability of the results of two PT-participants participating in two different PTs, even if no direct link has been established. Assessments like the one made here are also possible in practice, but it is not to be expected that it will happen on the scale as done in this project.

In reality, one may want judge whether the difference between two measurement results (on the same sample) from two laboratories participating in different PTs (if they participate in one PT, this case is covered by ISO 5725-2 [30]) can be explained from the results of participating in the respective PTs. Before modelling, some assumptions must be made:

1. The standard deviation obtained from the proficiency test is a good measure for the actual standard uncertainty of the measurement under study
2. Elementary requirements of metrological traceability are met
3. The standard deviations obtained in the proficiency tests are relevant for the range in which the results are to be compared

Most of the assumptions are self-explanatory and it is quite obvious why they are made. The third assumption is possible the most problematic one in practice. In many cases, the uncertainty associated with a measurement result is dependent on its value; it is in analytical chemistry often assumed that the relative uncertainty is more or less constant over a large range. Many performance characteristics (e.g., repeatability and reproducibility) in written standards indicate that this assumption is often, but certainly not always valid.

The first step in the comparison is to calculate the difference between the two measurement results x_1 and x_2

$$\Delta x_{12} = x_1 - x_2 \quad (11)$$

This difference is to be compared with the uncertainty associated with x_1 and x_2 respectively. From the proficiency tests in which the laboratories participate, there are standard deviations s_1 and s_2 , which can be assumed to be a "good measure" of the standard uncertainties u_1 and u_2 , associated with results x_1 and x_2 respectively. Consequently, the difference Δx_{12} is to be compared with

$$R_{12} = 2\sqrt{(u_1)^2 + (u_2)^2} \quad (12)$$

setting

$$u_1 = s_1 \text{ and } u_2 = s_2 \quad (13)$$

6.8 Results from the 5 PTs

The results from the 5 PTs generally indicate good comparability among participating laboratories in a PT, and across PTs. For the food sector as represented in this 2nd comparison, the assumptions made in 6.7 hold, which makes that in reality a comparison of results of laboratories *across* proficiency tests is possible, and the outcome is meaningful. One may argue that in practice it is not always possible to demonstrate on a case-by-case basis metrological traceability of the assigned values in the two PTs concerned, but –as appropriate– it is always possible to take appropriate measures when an actual sample is measured. The results of the PTs can aid interpreting the differences observed when two different laboratories measure the same sample.

7 Conclusions

The 2nd comparison in the CoEPT project has clearly demonstrated that it is possible to make useful inferences about results obtained on the same sample measured by two different laboratories. The implied assumptions concerning metrological traceability of the results, which is a prerequisite for being able to make a useful comparison, have proved to be valid for the PTs involved in the project. That this conclusion can be extended to other PTs is likely, but subject to proof.

Generally there was good agreement between any pair of assigned values in a dataset for a given parameter. From this agreement, it can be concluded that there is generally good agreement between the PTs of the participating PT-providers. Larger differences exist in the standard deviations and uncertainties estimated from the data reported by the participating laboratories, but they simply reflect differences in (average) performance of the participating laboratories. In conclusion, there is good comparability across the participating PTs for the characterisation of milk powder.

The biggest differences between the PT-providers are observed in the establishment of the standard deviation for performance rating and the evaluation of measurement uncertainty associated with the assigned value. The assigned values from the PT-providers agree usually with the reference value established for the certified reference material (CRM) used in this comparison. This fact not only leads to the conclusion that the approaches of the PT-providers for establishing an assigned value are valid, but also that the evaluation of uncertainty is at an acceptable level. Certainly the publication of ISO 13528 [15] is important to mention here, as it provides a reasonable approach to calculate the uncertainty associated with a consensus value.

A common protocol, based on ISO 13528 has been used to evaluate the datasets for all parameters of all PT-providers. This evaluation revealed also differences between the datasets, in particular in the level of scatter. It also made clear that the different approaches for establishing the standard deviation for performance rating come *in addition to* different levels of scatter.

8 References

- [1] International Organization for Standardization, “ISO/IEC Guide 43-1:1997, Proficiency testing by interlaboratory comparisons - Part 1: Development and operation of proficiency testing schemes”, ISO Geneva
- [2] Van der Veen A.M.H., Draft protocol 2nd comparison in CoEPT. Growth Project GTC1-2002-73002 CoEPT. Deliverable D15, NMi VSL, Delft (NL), **10 January 2005**
- [3] Van der Veen A.M.H., Robouch P., Örnemark U., “Comparison of PT protocols used in the food sector – Evaluation report of the 1st comparison within CoEPT, EU contract GTC1-2002-73002”, Delft, 30 March 2004
- [4] Örnemark U, Minutes from working group discussions, Geel (Belgium), 28 Oct. 2003, Protocol date: 2003-12-22
- [5] Örnemark U, Annex 1 to Agenda for sectoral discussions, 2nd CoEPT workshop, Geel (Belgium) 28 Oct. 2003
- [6] “CoEPT leaflet” (outlined by J van de Kreeke et al.), Feb. 2004
- [7] Örnemark U, Information and ordering of test samples for the 2nd comparison, 3 Feb. 2004
- [8] Örnemark U, Excel file *2004-08-29 2nd comp, Report Forms Food*, 30 Aug. 2004
- [9] Results report, Provider F02, 2005-09-29
- [10] Results report, Provider F01, 2005-09-29
- [11] Results report, Provider F03, 2005-12-09
- [12] Results report, Provider F05, 2005-12-16
- [13] Results report, Provider F04, 2005-11-24
- [14] Homogeneity test MUVA RM 56. Excel file *2004_HOM_RM56_Material 2nd study*, MUVA, 13 Aug. 2003
- [15] International Organization for Standardization, “ISO 13528:2005 Statistical methods for use in proficiency testing by interlaboratory comparisons”, ISO Geneva, 2005
- [16] Örnemark U, Minutes from working group discussions, Ede (The Netherlands), 8 Feb. 2005, Protocol date: 2005-02-27
- [17] International Organization for Standardization, “ISO 8968-1:2001 Milk - Determination of nitrogen content - Part 1: Kjeldahl method”, ISO Geneva, 2001
- [18] British Standard BS1741: section 5.2 - chemical analysis of liquid milk and cream
- [19] Provider F05 to U. Örnemark, personal communication 2005-01-13
- [20] International Organization for Standardization, “ISO 5537:2004. Dried milk - Determination of moisture content (Reference method)”, ISO, Geneva, 2004
- [21] Provider F04 to U. Örnemark, personal communication 2005-01-21
- [22] Provider F02 to U. Örnemark, personal communication 2005-01-13
- [23] Provider F01 to U. Örnemark, personal communication 2005-01-20
- [24] International Organization for Standardization, “ISO 1000:1992 - SI units and recommendations for the use of their multiples and of certain other units”, ISO, Geneva
- [25] International Organization for Standardization, “ISO 1736:2000 - Dried milk and dried milk products - Determination of fat content - Gravimetric method (Reference method)”, ISO Geneva, 2000

- [26] DIN 10477 - Determining the total ash of milk and milk products
- [27] International Organization for Standardization, "ISO 5765:2002. Dried milk, dried ice-mixes and processed cheese - Determination of lactose content. Part 1: Enzymatic method utilizing the glucose moiety of the lactose. Part 2: Enzymatic method utilizing the galactose moiety of the lactose. ISO Geneva, 2002
- [28] Regulation (EC) No 882/2004 of the European Parliament and the Council of the 29th April 2004 "on official controls performed to ensure the verification of compliance with feed and food law, animal health and animal welfare rules"
- [29] International Organization for Standardization, "ISO 5725-5:1998 Accuracy (trueness and precision) of measurement methods and results -- Part 5: Alternative methods for the determination of the precision of a standard measurement method", ISO Geneva, 1998
- [30] International Organization for Standardization, "ISO 5725-2:1994 Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method", Statistical methods for quality control, Vol. 2 (1994), pp 30-74
- [31] International Organization for Standardization, "ISO/IEC 17025:1999 General requirements for the competence of testing and calibration laboratories", ISO Geneva