ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Report on International Inter-laboratory Study on BMP tests

Hélène Fruteau de Laclos[1], Sasha Hafner[2], and Christof Holliger[3]

[1] Methaconsult, Préverenges, Switzerland, [2] Department of Engineering, Aarhus University, Denmark, [3] Laboratory for Environmental Biotechnology, EPFL, Lausanne, Switzerland

Lausanne, Switzerland

2 March 2018

Content

List of participating laboratories

Preliminary remark: The laboratory numbering in the results section does not correspond to the alphabetical order presented here. Thus, the anonymity of the laboratories in connection with their results should be guaranteed.

– Advanced Water Management Center, The University of Queensland, Brisbane, Australia;

– Bavarian State Research Center for Agriculture, Freising, Germany;

– Bioprocess Control AB, Lund, Sweden,

– Centre of Biological Engineering, University of Minho, Braga, Portugal;

– Chair of Urban Water Systems Engineering, Technical University of Munich, Garching, Germany,

– CRPA - Research Centre on Animal Production, Reggio Emilia, Italy

– DBFZ Deutsches Biomasse Forschungs-Zentrum, Leipzig, Germany,

– Delft University of Technology, Delft, The Netherlands,

– Department of Biotechnology University of Verona, Verona - Italy

– Department of Chemical Engineering, Institute of Technology, Universidade de Santiago de Compostela, Santiago de Compostela, Spain,

– Dept. of Chemical Engineering, Environmental Technology and Biotechnology, University of Southern Denmark, Odense M, Denmark.

– ENAC IIE Laboratory for Environmental Biotechnology, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland;

– Hamburg University of Technology (TUHH), Institute of Environmental Technology and Energy Economics Harburger Hamburg, Germany,

– Innotech Alberta, Hwy 16A & 75 Street, Vegreville, Alberta

– INRA, UR0050, Laboratoire de Biotechnologie de l'Environnement, Narbonne, France,

– INSA – DEEP, INSA de Lyon, Villeurbanne Cedex France

– Institut Polytechnique LaSalle Beauvais, Département des Sciences et Techniques Agro-Industrielles, Beauvais, France,

– Institute for Chemistry and Biotechnology, ZHAW School of Life Sciences and Facility Management, Wädenswil, Switzerland

– Instituto de la Grasa, Consejo Superior de Investigaciones Científicas, Seville, Spain,

– IRTA, Barcelona, Spain,

– LeAF, Wageningen, The Netherlands,

– LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France,

– National Research Council Canada, Montréal, Canada,

– OWS nv, Gent, Belgium.

– Politecnico di Milano-DICA, Milano, Italy,

– Sorbonne Universités, EA 4297 TIMR UTC/ESCOM, Compiègne, France,

– State Institute of Agricultural Engineering and Bioenergy, University of Hohenheim, Stuttgart, Germany

– Sub-Department of Environmental Technology, Wageningen University, Wageningen, The Netherlands,

– Technical University of Denmark, Lyngby, Denmark;

– The Swedish Centre for Resource Recovery, University of Borås, Borås, Sweden,

– University of Chemistry and Technology Prague, Prague, Czech Republic,

– University of Natural Resources and Life Sciences, Vienna, Austria,

– VERI (Veolia), Limay, France

List of abbreviations

BMP	Biochemical Methane Potential
VFA	Volatile Fatty Acids
COD	Chemical Oxygen Demand
TS	Total Solids
VS	Volatile Solids
ISR	Inoculum-Substrate Ratio
SD	Standard Deviation
RSD	Relative Standard Deviation
ADEME	Agence de la Maîtrise de l'Energie (FR)
VDLUFA	Verband Deutscher Landwirtschaflicher Unetrsuchnugs-unf Forschungsanstalten (DE)

1 Introduction

The biomethane potential (BMP) is a biological test used to determine the amount of methane that can be produced of a certain organic substrate. It is used in the following situations:

– The methane produced is the main, if not the only, source of revenues of anaerobic digestion (AD) plant. It is therefore essential to be able to predict beforehand the methane production from the organic substrates to be digested in the future AD plant and to establish an economic feasibility study of the project. Since the profitability is often very limited, a project can get rapidly non-profitable with only 15% less revenue.

– The contracts of the AD plant constructors include more and more guarantees on the performance regarding methane production that have to be between 85% to 95% of the substrate's BMP.

– Certain wastes that have a high BMP are no longer received for free but must be purchased, and it is important for AD plant planners and operators to evaluate the potential revenues in relation to the costs.

All these different utilizations of BMPs underline the importance of being capable to determine high reliability and precision this important parameter of the substrates to be digested. However, inter-laboratory studies of the past have shown that BMP test results vary considerably between laboratories, but they have not been able to clearly identify the parameters responsible for the observed variability.

During an international workshop in Leysin, Switzerland, in June 2015 we discussed intensively the problems with BMP tests and based on these discussions, we have published in WS&T a paper entitled « Towards a standardization of biomethane potential tests ». We also decided to carry out an inter-laboratory study based on the guidelines published in WS&T.

The outcome of this inter-laboratory study is the object of the present report.

2 Past inter-laboratory studies

Three inter-laboratory studies are summarized here in more detail in order to have a basis for comparison of the study carried in the framework of our project.

2.1 Inter-laboratory study of Raposo et al. 2011

In 2010 an inter-laboratory study organized with the support of the Spanish National Research Council and comparing BMP test results of solid and homogenous substrates from 20 European laboratories was published by F. Raposo et al.

2.1.1 Protocol

Four substrates have been sent to the participating laboratories :

- cellulose
- starch and gelatine
- two samples of mungo beans

No test conditions have been imposed except two parameters :

– all in triplicat

– Inoculum-Substrat ratios (ISR) fixed at

  - ISR=2 for starch and cellulose
  - ISR=3 for gelatine (to avoid potential ammoniac inhibition phenomena)
  - ISR = 1 et 2 for mungo beans

2.1.2 Results

BMP test results have been declared as outliers if they were <70% and >100% of the theoretical BMP calculated from the elementary composition of the substrates.
The results are summarized in Table 1.

Table 1. BMP test results of the European inter-laboratory study of Raposo et al. (2010)

	Starch	Cellulose	Gelatine	Mungo 2	Mungo 1
Theoretical BMP	414	414	433	434	434
All test results					
Mean (NmL/g VS)	320	340	300	340	330
Mini - maxi	126 - 417	175 - 412	124 - 480	189 - 447	170 - 437
Range (relative)	291 (90%)	237 (70%)	356 (119%)	258 (76%)	267 (81%)
Standard deviation	77 (24%)	52 (15%)	110 (37%)	63 (18%)	78 (24%)
Nb of outliers	4 / 17	3 / 17	9 / 17	5 / 17	
Without outliers					
Mean	350	350	380	370	370
Min. – max.	293 - 417	303 - 412	310 - 433	322 - 447	330 - 437
Range (relative)	124 (35%)	109 (31%)	123 (32%)	125 (34%)	107 (29%)
Standard deviation	33 (9%)	29 (8%)	42 (11%)	36 (10%)	35 (9%)

2.2 Inter-laboratory study of ADEME, France

Another inter-laboratory study has been carried out in France from 2012-2014 with the support of the “Agence de la Maîtrise de l’Energie” (ADEME). A total of 11 French laboratories participated.

2.2.1 Protocol

Four substrates have been sent to the participating laboratories:

- a mixture of proteins, starch, and fiber in its crude form
- the same mixture as above but dried and crushed (< 0.5 mm)
- straw
- mayonnaise

Two test series with two different protocols have been carried out. All has been done in triplicate. In each test series, two tests per lab have been done. In the first test series, each lab applied its own BMP test protocol, in the second, a partially harmonized protocol has been developed and applied by each laboratory:

- Addition of a nutrient solution and carbonate buffer
- A test of methanogenic activity with acetate has been done
- The Inoculum-Substrate ratio IRS = 2
- All measurements were done in triplicate
- Stop test when daily methane production is <1% of total accumulated gas

Four validation criteria have been introduced for the second test series:

- Relative standard deviation (RSD) of triplicates <10%
- pH at the end of the test > 6.5
- blanks should only produce <1/3 of total methane production
- methanogenic activity with acetate should be between 90-100% of the theoretical value

Outliers have only been removed after discussion with the laboratory that has done the test and only if a reasonable explanation could be provided, e.g. leak of the bottle concerned. The following criteria have been applied:

- If RSD <10% with two close values, the third one is eliminated
- If RSD >10% of three dispersed values, all three are eliminated

2.2.2 Results

The statistical analysis is based on ISO norms 13528 and 5725-1
The results after removing outliers are presented in Tables 2 and 3.

Table 2. Results of 1st test series of ADEME interlab study

	Crude mixture	Crushed mixture	Straw
Nb of values	50	68	59
Outliers	4	0	8
Mean (NmL/g VS)	425	403	267
Min.-max. (NmL/g VS)	289 - 629	250 - 481	175 - 370
Range (relative)	340 (80%)	231 (57%)	195 (73%)
Intra-laboratory repeatability	7%	4%	6%
Intra-laboratory reproducibility	9%	6%	8%
Inter-laboratory reproducibility	20%	17%	20%

Table 3. Results of 2nd test series of ADEME interlab study

	Crushed mixture	Straw	Mayonnaise
Nb values	69	53	56
Outliers	6	16	13
Mean (NmL/g VS)	405	277	848
Min. – max. (NmL/g VS)	260 - 525	195 - 370	660 - 1026
Range (relative)	265 (65%)	175 (63%)	366 (43%)
Intra-laboratory repeatability	4%	4%	4%
Intra-laboratory reproducibility	5%	7%	5%
Inter-laboratory reproducibility	19%	21%	13%

The tests were quite reproducible in one and the same laboratory.
A rather high number of outliers have been identified with straw and mayonnaise as substrates with very low and very high test results. The methanogenic activity acetate was not conclusive at all and has not been considered for test validation. The inter-laboratory reproducibility has not been improved by applying a more harmonized protocol.
The analysis of the parameters influencing the variability of test results showed in the 1st test series that the addition of a nutrient solution had a significant effect. In the 2nd test series, the methane measurement method using AMPTS II resulted in 15% lower BMPs compared with manual methods.

2.3 The inter-laboratory study of VDLUFA, Germany

For several years, VDLUFA has carried out inter-laboratory studies on parameters used to characterize substrates for AD, e.g. total solids, volatile solids, biogas production, and methane production. Here a brief summary is provided of the study carried out in 2015 with 25 participating laboratories.

2.3.1 Protocol

Three substrates were tested :

- Cellulose (positive control)
- Corn silage
- Wheat bran

Inoculum :

- 1% to 3% TS
- VS at least 50% of TS
- VFA < 500 mg/l HAc
- Preincubation of inoculum to decrease its methane production such that it produces <20% of the total methane production during the test

Substrate :

- Measurement of TS at 105°C and VS at 550°C

Procedure :

- All in triplicate
- ISR ≥ 2
- Purge with an inert gas (nitrogen or argon)
- Temperature 37°C ± 2°C
- Mesurement of final pH to verify that there was no acidification
- Mixing at least once a day

Test duration :

- Either minimum 25 days
- Or daily biogas production < 0.5% of total production for three consecutive days

Validation criteria :

- BMP of cellulose between 90% and 110% of reference value 745 NmL biogas/g VS, or 372.5 NmL methane for 50% CH4 in biogas → cellulose between 335 - 410 NmL CH4 /g VS

2.3.2 Results

Table 4. Results of VDLUFA inter-laboratory study of 2015

	Cellulose	Corn silage	Bran
P laboratories	26	27	26
P1 validated laboratories	20	23	22
N all test results	85	90	87
N1 validated results	66	77	74
Mean (NmL/g VS)	368	353	360
Tolerance (moy ± 2 S_R)	318 – 417	265 - 440	286 - 435
RSD of repeatability	3.0%	4.1%	3.9%
RSD of reproducibility	6.7%	12.4%	10.3%
all test results			
Min.- max. (NmL/g VS)	243 - 406	249 – 427	281 – 425
Range (relative)	163 (44%)	178 (50%)	144 (40%)
without outliers			
Min.- max. (NmL/g VS)	331 - 406	249 – 427	292 - 425
Range (relative)	75 (20%)	178 (50%)	133 (37%)

Based on cellulose BMP data, 6 out of 25 laboratories have been removed as outliers.

The RSDs of repeatability and reproducibility were rather low but the ranges were still quite high.

2.4 Conclusions

Although the RSD of repeatability and reproducibility of two of three inter-laboratory studies were quite low, the range of the test results remains rather high. Although removing outliers helped to decrease these ranges, the criteria for outliers cannot be applied if one sends substrates to a single laboratory. Hence, the risk of inaccurate BMP values if sending substrates to a single laboratory is still unacceptably high.

3 Setup of present inter-laboratory study

3.1 Organisation

Thirty-three laboratories volunteered to participate in this inter-laboratory study (Annex 1).

Two test series have been carried out with interval of 1-2 months. Four laboratories have only carried out one test.

Three substrates, cellulose, a trace element solution, and a vitamin solution has been sent to each laboratory as well as an Excel file for data reporting.

Despite many reminders, the test results were sent back often very late.

3.2 The substrates

Three substrates have been chosen based on the following criteria :

- high stability to permit long distance transport by mail, hence as dry as possible
- Homogeneity : if possible particle size already < 1 mm
- different compositions

CARGILL that produces different animal feeds has provided for free three products that we used in the inter-laboratory study. An analysis of the elementary composition has been carried out in order to be able to calculate the theoretical BMP using the Buswell equation.

Substrat A *Substrat C* *Substrat B*

a. Substrat A = Pig feed

Substrate A contains 30% wheat, 19% triticale, 15% barley, 8% protein peas, 6.7% bran, 5% rapeseed cake, 4.9% soya cake, 1.2% fat and the rest are minerals, amino acids, and oligo elements.

- Elemental composition : C17H31O13N
- Theoretical BMP : 428 NmL CH4/g VS

b. Substrat B = Fodder flour

Substrate B contains 44 % cellulose, 37.8% starch and 15% nitrogen-containing compounds.

- Elemental composition : C23H38O16N
- Theoretical BMP : 454 NmL CH4/g VS

c. Substrat C = Fodder called Probos

Substrate C contains 19% cod liver oil, 14% sunflower seed cake, 12% rapeseed cake, 12% soya cake, 10% flaxseed cake, 4% Ca_2PO_4, 3% wheat germs, 2.8% NaCl, 2% oat, 1% yeast, and the rest is fodder flour and bran.

- Elemental composition : C18H32O8N
- Theoretical BMP : 608 NmL CH4/g VS

3.3 The Protocol

The protocol was based on the WS&T paper entitled « Towards a standardization of biomethane potential tests ». Some parameters have been imposed. However, not all laboratories followed the protocol as requested.

3.3.1 Inoculum

The inoculum should be taken from an anaerobic digester with stable operation, fed with a complex substrate or a mixtures of substrate in order to contain a diverse microbial community. By preference, the inoculum should be taken from anaerobic digester of a wastewater treatment plant or an agricultural plant digesting manure.

The inoculum should undergo the least treatment as possible, e.g. no sieving if possible. The analysis of TS and VS in triplicate is compulsory and the VS content should be between 15 and 40 g/l.

The inoculum should have the following characteristics:

- pH between 7.0 and 8.5
- VFA < 1 g/l HAc
- NH4 < 2.5 g/l N-NH4
- Alkalinity > 3 g/l $CaCO_3$

The inoculum should only be stored for a short period (2-5 days) at ambient temperature or at the temperature at which the BMP test will be carried out.

The inoculum should an endogenous methane production that is ≤ 20% of the total production (inoculum+substrate).

3.3.2 Substrates

The analysis of TS and VS in triplicate is compulsory, the analysis of COD is optional. The substrates can be stored at 4-15°C in the dark and at a dry place.

3.3.3 Set-up

All tests and analyses have to be carried out in triplicate

Each test bottle should contain at least 2 g of substrate.

The trace element and vitamin solution have to be added at 1ml/l each. If alkalinity is < 3 g/l $CaCO_3$, bicarbonate has to be added to reach at least 3 g/l $CaCO_3$.

The purge gas should be by preference N_2/CO_2 with 20-40% CO_2 but 100% N_2 is acceptable

The total VS concentration should between 20 – 60 g/l.

The ISR has to be 2 for substrates A and B, and 4 for substrate C.

The incubation temperature has to be 37°C ± 2°C. Mixing is compulsory, at least once a day manually.

There are no restrictions on the way methane production is measured but if the method requires measurement of gas composition, this has to be done at each measuring point. In addition, the ambient temperature and atmospheric pressure has to be recorded. If an AMPTS II is used, one has to inactivate the option “eliminate overestimation”, the bottles have to be filled with 450 ml (total volume) and purged with a mix of N_2/CO_2.

The test duration is terminated at the point where daily methane production is <1% of the net cumulated methane production (substrate – blank) during 3 consecutive days.

3.3.4 Reporting of test results

The volume of methane produced has to be expressed as dry gas under standard conditions. The total production of one bottle has to be provided without subtracting the blank or already calculating the volume of methane produced per g VS of substrate.

3.3.5 Method of calculation

– Net methane production per batch = (total production of a batch) – (mean of the production of the 3 blanks)
If the same amount of inoculum was not used in each bottle, the net methane production has be corrected accordingly.

- BMP per batch = net methane production / VS mass of substrate
- Mean of the BMPs of the 3 replicates for a test and a substrate
- The standard deviation is calculated as follows :

$SD_{BMP} = \sqrt{SD^2_{blanc} + SD^2_{substrat/control}}$ with

(1) SD^2_{blanc} = Variance of the methane production of the blanks

- (2) $SD^2_{substrat/control}$ = Variance of the methane production of the substrates

For the statistical calculations, the following applies :

- o BMP of one batch = x_i
- o Number of replicates = n (=3)
- o 1 cell = the (3) results of the replicates
- o The mean of the BMPs of one cell = $\bar{x}$
- o The standard deviation of the BMPs of one cell = s
- o Number de laboratories = p

– Estimation of the mean of the BMPs of one cell

$\bar{x} = \sum_1^n x/n$

– Estimation of the standard deviation of the BMPs of one cell

$s = \sqrt{\sum_1^n (x - \overline{x})^2/(n-1)}$

– Estimation of repeatability standard deviation s_r

In this study two series of tests have been carried out. To calculate s_r, all the results of p laboratories for the two series have been considered:

$s_r = \sqrt{\sum_1^{p'} s^2 /p'}$

where p'= total number of independent tests, about 2p

– Estimation of reproducibility standard deviation s_R

$s_R = \sqrt{s_L^2 + s_r^2}$

– Estimation of standard deviation of cell averages $s_{\bar{x}}$

$s_{\bar{x}} = \sqrt{\sum_1^{p'}(\bar{x} - \bar{\bar{x}})^2 / (p' - 1)}$

where $\bar{\bar{x}} = \sum_1^{p'} \bar{x} / p'$

– Estimation of between laboratories variance s_L^2

$s_L^2 = s_{\bar{x}}^2 - s_r^2 / n$

– Estimation of standard deviation of "intra-laboratory reproducibility s_{Rs}"

This is the standard deviation of reproducibility between two test series by the same laboratory that are however not strictly under repeatability conditions.

$s_{Rs} = \sqrt{s_m^2 + s_r^2}$

– Estimation of the variance between test series s_m^2

$s_m^2 = s_{s1-2}^2 - s_r^2 / n$

– Estimation of standard deviation of means of series (simplified calculation for 2 series)

$s_{s1-2} = (\bar{x}_{s1} - \bar{x}_{s2}) / \sqrt{2}$

where $\bar{x}_{s1}$ et $\bar{x}_{s2}$ are the means of cell of series 1 and 2 per laboratory, respectively.

4 Results

4.1 Checking and correcting submitted data

Despite providing a specific Excel file with clear indications how the data should be submitted, several laboratories have chosen to modify the file and provide the data in a different format or already treated. It was cumbersome and time consuming to check all the files and get them uniform.

Many laboratories did not determine TS and VS in triplicate, neither for the substrates and cellulose nor for the inoculum.

4.2 Statistical analysis according the norms for inter-laboratory studies

4.2.1 Measurement of TS and VS

The means of TS and VS measurements of all 33 laboratories are summarized in Table 5 and individual measurements are shown in Figure 1. Although the SDs are quite low, there was surprisingly high variability in these measurements (Fig. 1). For cellulose, VS of cellulose varied from 89-96%, for the substrates VS varied from 77-88%. In most cases, differences among laboratories were small, but for each substrate, some laboratories (2-6) reported unusually low values (Fig. 1). These differences finally also can contribute significantly to the variation in BMPs among the different laboratories.

Table 5. : TS and VS of cellulose and substrates A, B and C

	cellulose	SA	SB	SC
TS mean ± SD (%)	94.94 ± 0.79%	88.81 ± 0.64%	89.04 ± 0.63%	92.28 ± 0.76%
VS/TS mean ± SD (%)	99.82 ± 1.00%	91.89 ± 4.33%	95.81 ± 3.90%	87.08 ± 2.60%
VS mean ± SD (%)	94.68 ± 1.38%	81.61 ± 4.30%	85.82 ± 3.95%	80.35 ± 2.56%

Although initially not considered as parameter that can influence the outcome of a BMP test, this should be reconsidered and discussed. It has to be decided on criteria for validation of these measurements, e.g. all in triplicate which many laboratories didn't do, and including the measurement error in SD calculation.

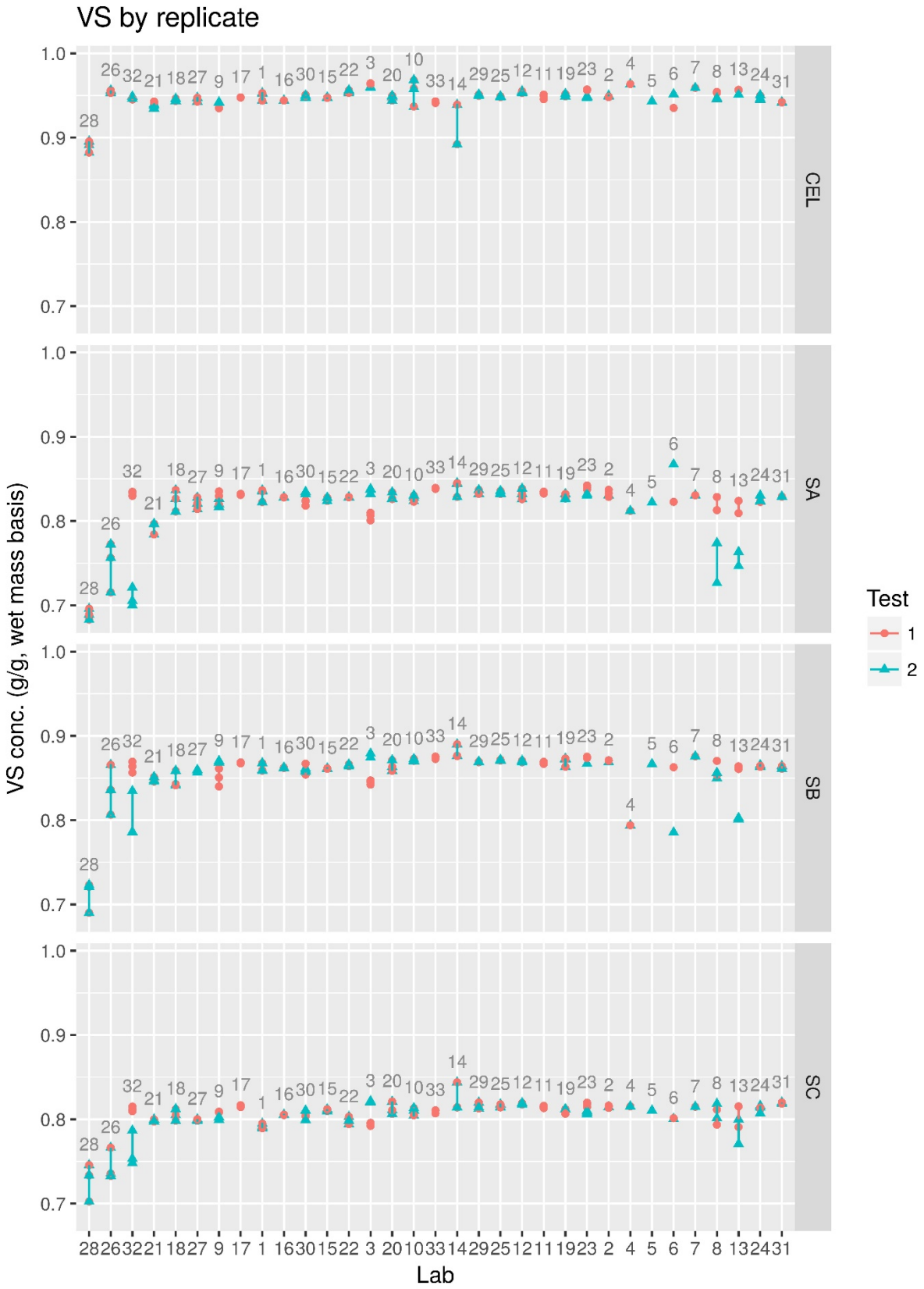

Figure 1. Variation of VS content determined for cellulose and substrates A, B, and C

4.2.2 Global BMP test results

All test results per batch, per lab, and per test series are depicted in Figure 2

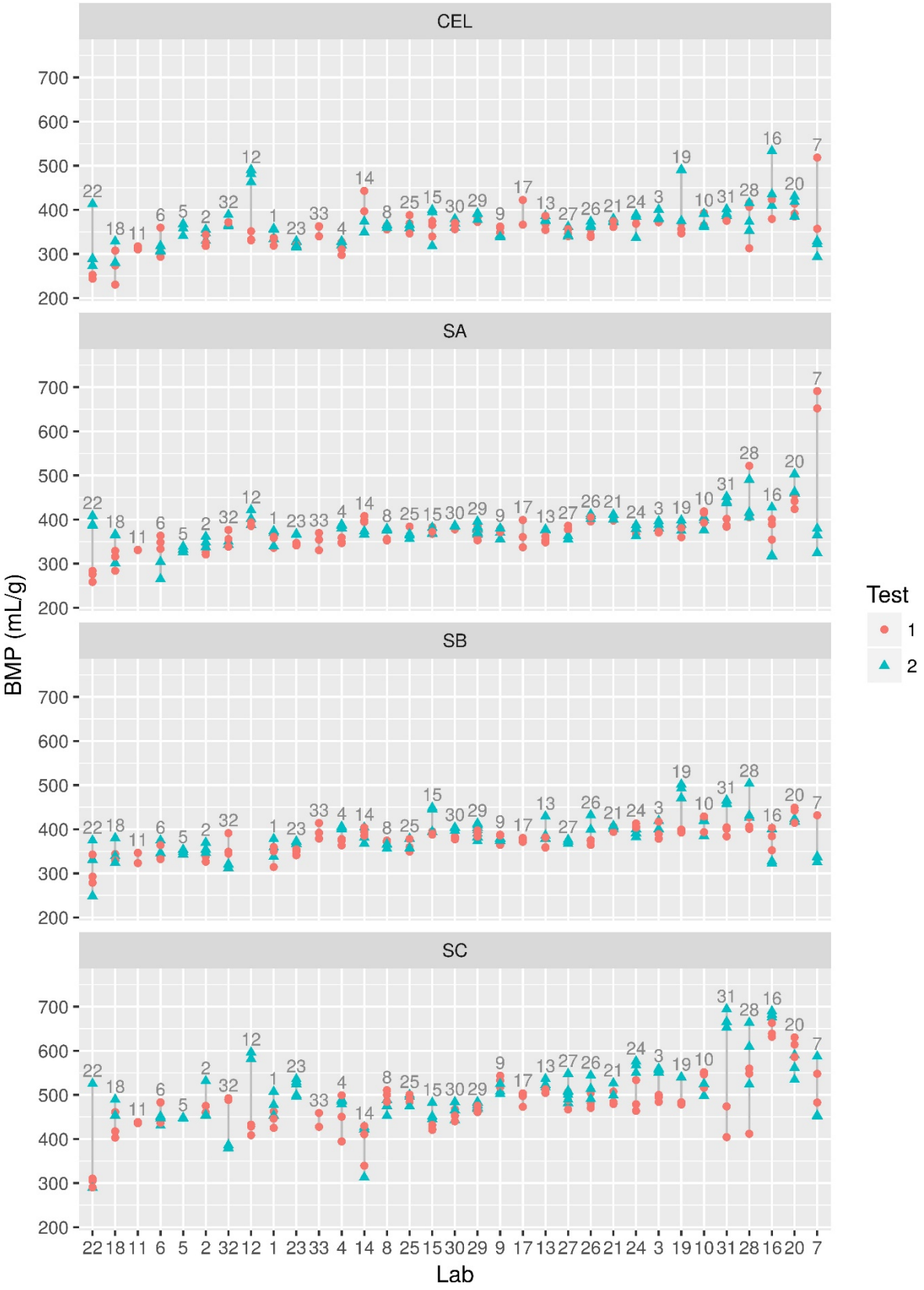

Figure 2. BMP per batch, per lab and per test series for cellulose and substrates A, B, and C. Three observations from lab 12 below 0 for substrate B, and four observations from lab 7 above 750 NmL/g VS are not shown.

The statistical parameters for all substrates are summarized in Table 7.

Table 7. Statistical parameters of the inter-laboratory study

	cellulose	SA	SB	SC
Number of laboratories	33	33	33	33
Number of test series [1]	2	2	2	2
Number of replicates	3	3	3	3
Number of calculated means	63	62	59	62
Theoretical BMP (Nl/kg VS)	414	428	457	608
Mean measured BMP (NmL/g VS)	365.3	380.8	388.9	494.5
Median measured BMP (NmL/g VS)	367.9	376.2	380.3	490.4
Min. – max. (NmL/g VS)	236.8 668.8	272.7 717.7	304.7 691.7	301.9 683.1
Range (NmL/g VS) (relative % in parentheses)	432.1 (118%)	445.0 (117%)	387.0 (100%)	381.2 (77%)
Robust mean [2] (NmL/g VS)	362.8	375.6	383.2	490.4
Robust SD [2] (NmL/g VS)	27.6	27.7	34.4	44.0
Intra-laboratory repeatability [3]	**8.8%**	**7.0%**	**6.9%**	**9.2%**
Intra-laboratory reproducibility [4]	**10.5%**	**8.4%**	**8.1%**	**9.5%**
Inter-laboratory reproducibility	**17.1%**	**15.8%**	**15.1%**	**15.3%**

[1] Three laboratories provided data for only 1 test
[2] Mean and SD calculated with algorithm A of the ISO norm 5725 which allows to remove extreme values without prior statistical tests
[3] Tests with only 1 or 2 replicates were removed
[4] The labs with only 1 test series are not included

The robust means of cellulose and the three substrates accounted for 80.6-87.6% of the theoretical BMP with the lowest percentage for substrate C and the highest for cellulose and substrate A. The robust mean of cellulose is very similar to the one calculated in VDLUFA inter-laboratory studies. This indicates that this value is perhaps a good reference to define new validation criteria for the positive control. The lower limit of cellulose that would allow validation according to our guideline is very close to this value (352 ml vs. 363 ml). This lower limit should therefore perhaps by modified (see below).

It was surprising to see that the range was high for all substrates and the largest for cellulose which is normally a very homogeneous substrate that should not result in such high variance of BMP results.

The RSD of intra-laboratory repeatability is higher compared to the ADEME and VDLUFA studies. One should note however that for the calculations done here only one test with extreme values has been removed together with values for substrate B of one lab that observed inhibition effects with no methane production with this substrate, while in the other studies several outliers were removed before calculating repeatability.

Inter-laboratory reproducibility lays in between the ADEME and VDLUFA studies, hence it is comparable and also not satisfactory.

Nevertheless, although these results describe the dispersion of BMP test results for the whole set of participating laboratories, they do not allow to achieve the desired goal to render the result of a measurement carried out in any laboratory reliable. The range of the test results are simply too high.

4.3 Application of original validation criteria

As indicated in the introduction, BMP measurements should be sufficiently reliable to allow use for economic or even contractual purposes. Besides defining a test protocol with clear requirements for all steps of the method, the use of validation criteria is another way to make the results more reliable.

4.3.1 Original validation criteria

In the guidelines published in WS&T the following validation criteria have been defined:

– Criterion 1 : RSD of the blanks triplicates ≤ 5%
– Criterion 2 : RSD of cellulose and homogenous substrate triplicates ≤ 5%
– Criterion 3 : BMP of cellulose is between 352 NmL/g VS and 414 NmL/g VS (85-100% of theoretical value)

The SD for the BMPs has to be calculated as follows:

$$SD_{BMP} = \sqrt{SD_{blank}^2 + SD_{substrate/control}^2}$$

4.3.2 Application of validation criteria for cellulose

The outcome of the application of the original validation criteria is summarized in Table 8. Application of criteria 1, 2, and 3 separately eliminated about the same number of tests (14-15). When applying criteria 1 and 2 together, less than half of the tests could be validated and the range was still rather high. Applying criterion 3 in addition, obviously decreased the range and only 19 tests (30%) could finally be validated.

Tableau 8: Application of original validation criteria on cellulose test data

Application of criteria	Nb of tests validated	Min. –max.	Range	SD
none	63	237 - 669	432	56
Criterion 1	39	237 - 669	432	66
Criterion 2	38	295 - 479	183	29
Criteria 1+2	28	312 - 479	167	30
Criterion 3	38	354 - 413	59	14
Criteria 1+2+3	19	354 - 379	25	8.3

Figure 3 shows the final result for cellulose and the three substrates.

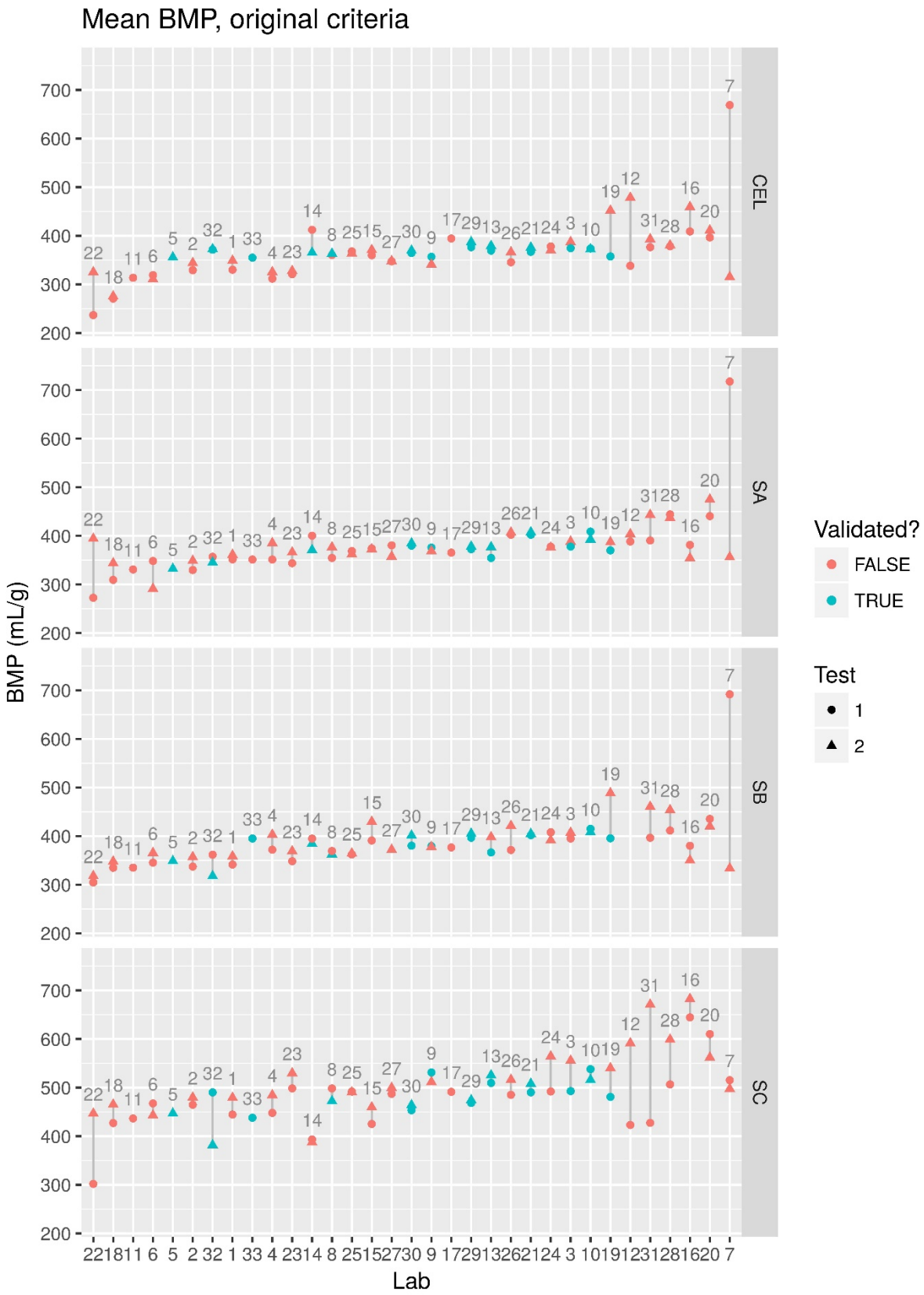

Figure 3. Application of original validation criteria for cellulose and substrates A, B, and C. Blue symbols: validated tests, red symbols: non-validated tests.

4.3.3 Application of validation criteria for substrate A

The results of tests for which the cellulose BMP could not be validated have been removed. For the remaining BMPs of substrate A, criterion 2 has been applied. When applying criterion 2 alone, 22 tests could not be validated. When applying criterion 2 to the test results for substrate A that were validated based on criteria 1-3 for cellulose BMPs, additional 3 tests could not be validated. Hence, at the end, only 16 out of 62 tests (26%) could be validated for substrate A.

Table 9. Application of original validation criteria on substrate A test data

Application of criteria	Nb of tests validated	Min. – max.	Range	SD[1]
none	62	273 - 718	445	55
Criteria 1-3 to cellulose	19	332 - 409	76	21
Criterion 2 to substrate A	40	330 - 443	114	23
All criteria	16	332 - 409	76	21

[1] Standard deviation among laboratories

4.3.4 Application of validation criteria for substrate B

The results of tests for which the cellulose BMP could not be validated have been removed. In addition, two test results where inhibition has been observed have been removed as well before applying criterion 2 for the remaining BMPs of substrate B. Here, also 22 tests could not be validated when applying criterion 2 alone, and 4 additional ones when applying criterion 2 to tests validated for cellulose. Finally, only 16 out of 59 tests (25%) could be validated.

Table 10. Application of original validation criteria on substrate B test data

Application of criteria	Nb of tests validated	Min. – max.	Range	SD
none	59	305 - 692	387	54
Criteria 1-3 to cellulose	19	318 - 440	122	27
Criterion 2 to substrate B	37	318 - 488	170	33
All criteria	16	318 - 415	97	26

4.3.5 Application of validation criteria for substrate C

The results of tests for which the cellulose BMP could not be validated have been removed. For the remaining BMPs of substrate C, criterion 2 has been applied. Here, 23 tests could not be validated when applying criterion 2 alone, and 2 additional ones when applying criterion 2 to tests validated for cellulose. Finally, only 18 out of 62 tests (27%) could be validated.

Table 11. Application of original validation criteria on substrate C test data

Applied criteria	Nb of tests validated	Min. – max.	Range	SD
none	62	302 - 683	382	65
Criteria 1-3 to cellulose	19	381 - 538	156	44
Criterion 2 to substrate C	39	302 - 683	382	64
All criteria	18	381 - 538	156	40

4.3.6 Discussion of the original validation criteria

The application of the original validation criteria shows that only about a quarter of the tests could be validated. This indicates that the criteria are perhaps somewhat too restrictive.

With regard to criterion 1 for example, the RSD becomes rapidly high if the residual methane production of the inoculum is low. This penalizes the laboratories that comply with the recommendation on a maximum production of 20% by the blanks. Moreover, the variability of the methane production of the blanks is taken into account in the calculation of the SD of the BMP as proposed above (and generally makes a small contribution), and the application of criterion 1 did not decrease the range of the BMPs measured for cellulose. It is therefore proposed to exclude this criterion.

The criterion 2 (RSD ≤5%) is fulfilled by only 38 out of 63 tests for cellulose. Here too, one could think about loosen the criterion and to validate BMP test results with an RSD ≤10%.

Criterion 3 concerns the acceptable range of the positive control BMP for which a theoretical BMP can be calculated. The robust mean for cellulose was 363 Nl/kg VS and the lower limit of the acceptable range initially proposed is 352 Nl/kg VS. This small difference (compared to variation among laboratories) suggests that this acceptable lower limit should be lowered. It is proposed to decrease this limit to 331 Nl/kg VS which is 80% of the theoretical limit and about 90% of the observed robust mean. One could perhaps also reconsider the upper limit. At present 100% of the theoretical limit, one could decrease it to 110% of the robust mean, e.g. to 401 NmL/g VS which is 97% of the theoretical limit.

An evaluation of potential associations between passing the individual criteria showed that tests that met the substrate RSD criterion were more likely to meet the cellulose BMP criterion. There was no association between inoculum RSD and cellulose BMP but there is a strong association between passing the inoculum RSD criterion and the substrate RSD criterion which is perhaps not surprising.

4.4 Application of revised validation criteria

4.4.1 Proposed revised criteria

– Criterion 1 : abolish

– Criterion 2 : RSD of cellulose and homogenous substrate triplicates ≤ 10%

– Criterion 3 : BMP of cellulose comprised between 331 Nl/gVS and 414 Nl/gVS

4.4.2 Application of revised validation criteria for cellulose

The results of the application of the revised validation criteria are summarized in Table 12. Application of criteria 2 and 3 separately eliminated 13 and 19 tests, respectively. When applying criteria 2 and 3 together, 40 tests (63%) could be validated which is more than twice the number of validated tests with the original validation criteria.

Table 12. Application of revised criteria to cellulose results

Applied criteria	Nb of tests validated	Min. – max.	Range	SD
none	63	237 - 669	432	56
Criterion 2	50	237 - 479	242	35
Criterion 3	44	339 - 413	74	17.0
Criteria 2+3	40	339 - 413	74	16.7

Figure 4 shows the final result for cellulose and the three substrates.

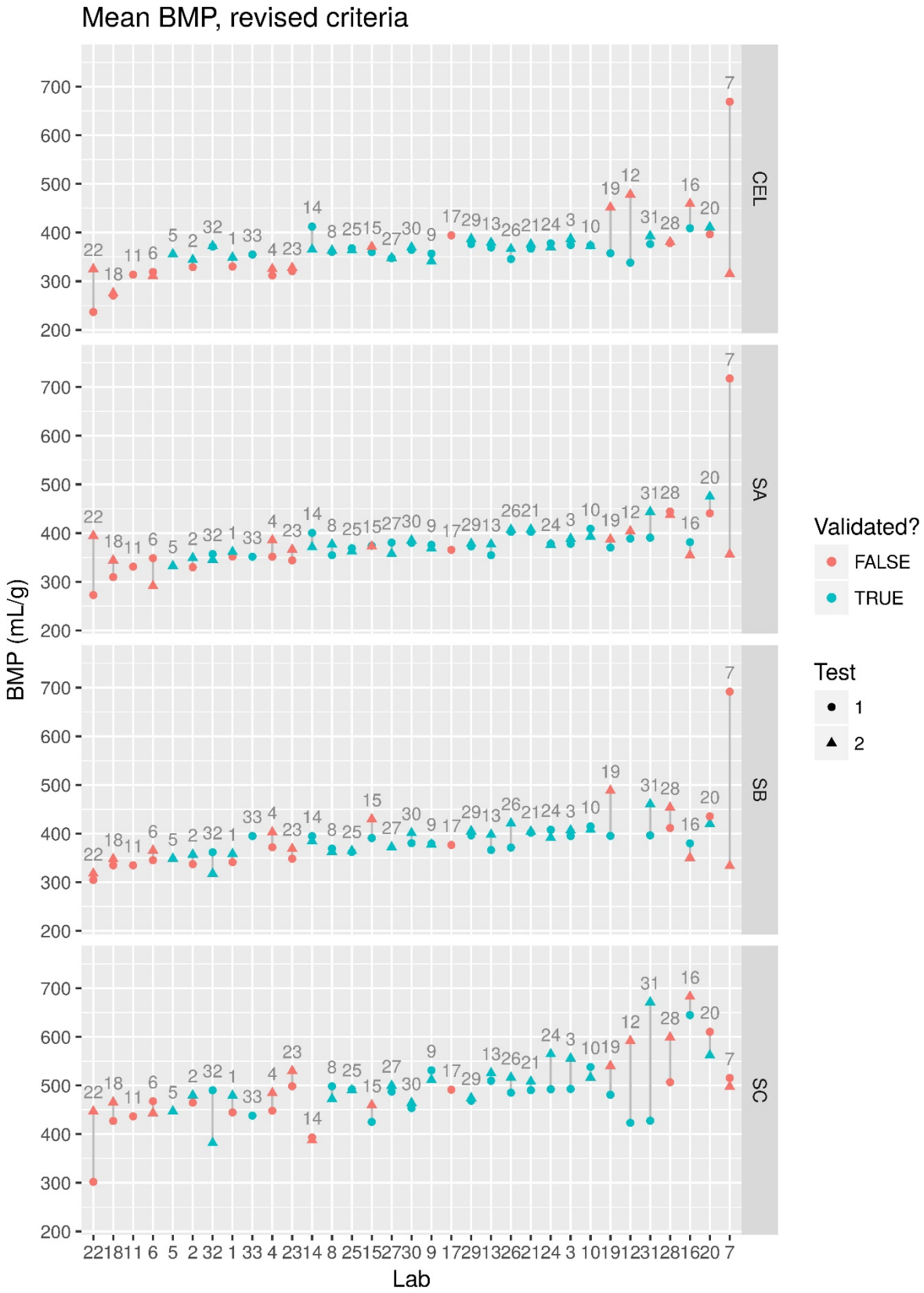

Figure 4. Application of revised validation criteria for cellulose and substrates A, B, and C.
Blue symbols: validated tests, red symbols: non-validated tests.

4.4.3 Application of revised validation criteria for substrate A

The results of tests for which the cellulose BMP could not be validated have been removed. For the remaining BMPs of substrate A, criterion 2 has been applied. When applying criterion 2 alone, only 10 tests could not be validated. When applying criterion 2 to the test results for substrate A that were validated based on revised criteria 2+3 for cellulose BMPs, all tests could be validated. Hence, at the end, 40 out of 62 tests (65%) could be validated for substrate A.

Table 13. Application of revised criteria to substrate A

Applied criteria	Nb of tests validated	Min. – max.	Range	SD
none	62	273 - 718	445	55
Criteria 2+3 to cellulose	40	332 - 444	111	21
Criterion 2 to substrate A	52	273 - 478	205	31
All criteria	40	332 - 444	111	21

4.4.4 Application of revised validation criteria for substrate B

When applying criterion 2 alone, only 8 tests could not be validated. When applying criterion 2 to the test results for substrate B that were validated based on revised criteria 2+3 for cellulose BMPs, all tests could be validated. Hence, at the end, 37 out of 62 tests (60%) could be validated for substrate B.

Table 14. Application of revised criteria to substrate B

Applied criteria	Nb of tests validated	Min. – max.	Range	SD
none	59	305 - 692	387	54
Criteria 2+3 to cellulose	37	318 - 460	142	27
Criterion 2 to substrate B	51	318 - 488	170	32
All criteria	37	318 - 460	142	26

4.4.5 Application of revised validation criteria for substrate C

When applying criterion 2 alone, only 11 tests could not be validated. When applying criterion 2 to the test results for substrate C that were validated based on revised criteria 2+3 for cellulose BMPs, one additional test could not be validated. Hence, at the end, 39 out of 62 tests (63%) could be validated for substrate C.

Table 15. Application of revised criteria to substrate C

Applied criteria	Nb of tests validated	Min. – max.	Range	SD
none	62	302 - 683	381	65
Criteria 2+3 to cellulose	40	382 - 671	289	58
Criterion 2 to substrate C	51	302 - 683	381	64
All criteria	39	382 - 671	289	54

4.4.6 Impact of application of original and revised criteria

The impact of validation criteria on the number of validated tests, the range of BMP test results and the RSD for inter-laboratory reproducibility are summarized in Table 16. Figure 5 shows the same analysis but with box plots.

Tableau 16: Impact of application of original and revised validation criteria

Applied criteria	Nb of tests validated	Min. – max.	Range	RSD reproducibility[1]
Substrate A				
none	62	273 - 748	445	15.8%
original criteria	16	332 - 409	76	
revised criteria	39	332 - 444	111	8.1%
Substrate B				
none	59	305 - 692	387	15.1%
original criteria	15	318 - 415	97	
revised criteria	36	318 - 460	142	8.8%
Substrate C				
none	62	302 - 683	381	15.3%
original criteria	17	381 - 538	156	
revised criteria	37	382 - 671	289	12.7%

[1] RSD of reproducibility has not been calculated for application of original criteria.

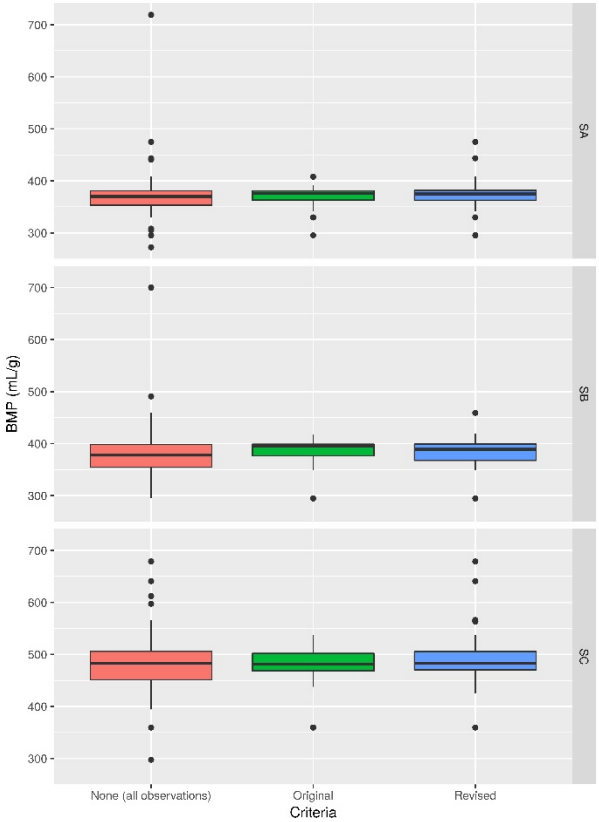

Figure 5. Box plots illustrating the impact of validation criteria application. The heavy line shows the median, the box shows the 25th and 75th percentiles, extreme observations (more than 1.5 times the interquartile range beyond the box) are plotted as points, and whiskers show the extent of other observations.

The application of validation criteria allowed reducing considerably the range of measured BMPs. This is especially true for the original criteria; however these criteria also only allowed validating about 25% of the tests. The revised criteria allowed to validate about 60% of the tests but considerably increased the range of measured BMPs. Whereas the range for substrates A and B was a factor of 1.46 larger, it was 1.85 larger for substrate C.

It is important to note that in the present study no outlier based on too high deviation from the mean has been removed. Only validation criteria have been applied that can be applied by individual laboratories since finally BMPs are determined by a single laboratory and not a set of laboratories as in inter-laboratory studies.

The application of validation criteria in addition allowed to decrease the RSD of inter-laboratory reproducibility except for substrate C where the decrease is only marginal.

4.5 Test validation per laboratory

This evaluation is limited to the BMP test results for cellulose. Table 17 shows that almost two third of the laboratories had no test validated when the original criteria were applied. This decreased to one third with the revised criteria.

Table 17. Number of validated tests per laboratory

	original criteria	**revised criteria**
Laboratories with all tests validated (1/1 ou 2/2)	7	18
Laboratories with one validated test out of two	7	5
Laboratories with no validated test	**19**	**10**

The results also show that in the majority of cases (26 or 28 out of 33) either all tests of a laboratory are validated or none which suggests that the measurement error is not random but related to the method used by the laboratory. The method used also includes the inoculum used.

4.6 Difference between the two tests

Figure 6 illustrates the difference between the two test series. Although there is a tendency that laboratories with a low difference have test results that could be validated, there are also laboratories with large differences between the tests and still both tests were validated, e.g labs 32 and 31. Analysis using mixed-effects models showed strong evidence overall of test biases (Section 4.8).

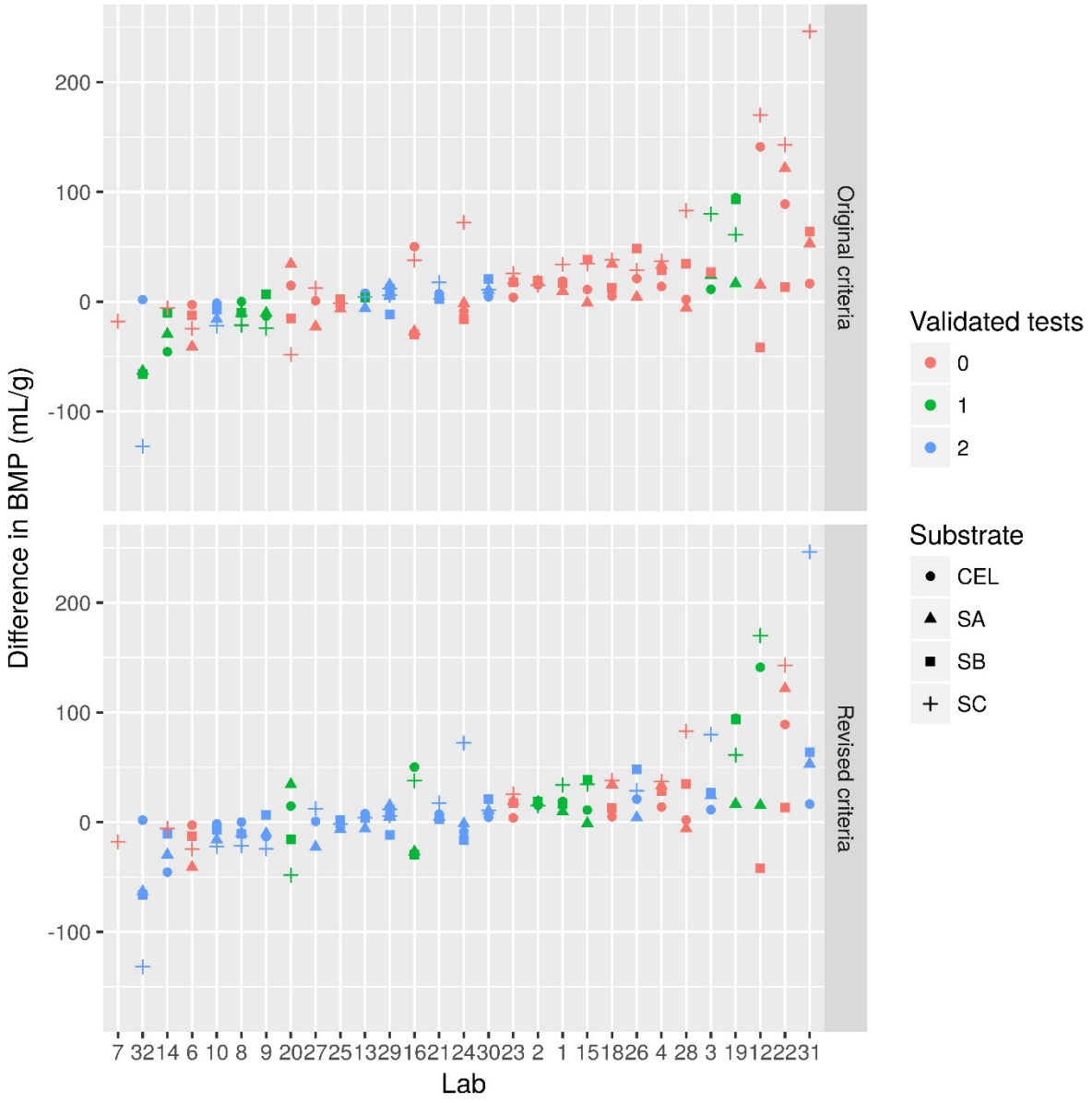

Figure 6. Difference between the two tests for each laboratory.

4.7 Application of test protocol recommendations

4.7.1 Inoculum

		Nb of laboratories
Origin	AD at WWTP	16
	AD of manure and co-substrates	6
	AD of manure only	5
	UASB reactor	3
	Laboratory AD batch	2
	Thermophilic AD plug-flow reactor	1
VS concentration	< 10 g/l	2
	> 50 g/l	3
	10 - 15 ou 40 - 50 g/l	4
	15 - 40 g/l	24
Inoculum quality criteria		too few data available

The majority of laboratories has used an inoculum from a WWTP, as recommended, or digested manure. The majority also followed the recommendations for the VS concentration. Only very few laboratories verified the quality criteria for inocula and therefore no conclusions can be drawn for these parameters.

4.7.2 Methane production by the blanks

Methane should have been <20% of the total production (blank + cellulose). However, only in about a quarter of the tests this recommendation has been fulfilled. In the majority of tests (44 out of 63) the production of the blanks was <30% and there was no correlation between the production of the blanks and the validation of the tests. This suggests that this recommendation should be revised.

		Nb of tests
Production	< 20%	16
	20 -30%	28
	30 - 40%	11
	< 40%	8

4.7.3 Minimum mass of substrate added per batch

		Nb of laboratories
Mass of substrate/cellulose added	> 2g	28
	1 - 2 g	2
	< 1g	3

Not all laboratories have followed the recommendation of adding at least 2 g VS of substrate or cellulose. However, there was no correlation between the respect of this recommendation and the validation of the tests.

4.7.4 Test duration

During the study, different problems have been encountered regarding the test duration criterion. The test should be stopped when the daily net methane production is <1% of the total net methane production and that during three consecutive days. Certain laboratories did not stop the test at this moment and continued until no methane was produced anymore. The difference between the 1%-criterion and the ultimate methane production was up to 5-6%. This suggests that the test duration criterion should be revisited.
Another problem encountered is the fact that in a test with substrates that have different methane production kinetics, the test duration might be different for the different substrates. If that is the case, the production of the blank has to be recorded at the different moments when methane production on a specific substrate has reached the test duration criterion and not only when the last one has reached the end of test duration. This has to be described more clearly in the guidelines.

4.8 Laboratory versus test biases

Mixed effect models were used to quantify variability among laboratories and between the two tests carried out by each laboratory. Observations for all analyses described here are bottle-level BMP estimates made with fixed VS concentrations (i.e., a single mean VS concentration for each substrate calculated from reported values after removing unusual values). This differs from the data presented in other sections where BMPs were calculated with the VS concentrations measured by each laboratory.

Not surprisingly, results showed that there were clear laboratory biases, i.e., considering all observations for all substrates, individual laboratories had a tendency to produce BMP estimates above or below the

overall mean response (P < 0.0001). (All P values presented in this section are based on likelihood ratio tests.) Laboratory biases can be seen most easily in Figure 3 where results from the different substrates generally show a similar position relative to results from other laboratories. Furthermore, the two tests within individual laboratories were in many cases close to each other compared to differences among laboratories. However, test biases were also present (P < 0.0001), meaning that results from an individual test performed in one particular laboratory were, on average, noticeably smaller or larger than the mean laboratory result. Test biases can be clearly seen in Fig. 6 where results from the four substrates are generally grouped together.

The presence of biases is not surprising, and it is their magnitude that determines how important they are. Variation in the magnitude of these biases depended on substrate (P < 0.03), with the highest values associated with substrate C, and lowest values with substrate A (Table 18). This difference is consistent with overall variability estimates (Table 7). Estimated values of laboratory and test standard deviation ranged from just below 20 ml/g VS for the substrate A test effect to more than 50 ml/g VS for the substrate C test effect (Table 18). The magnitude of these sources of variation was significant compared to within-lab, within-test (residual) variation (standard deviation of 24 ml/g VS). Both of these types of biases clearly make a significant contribution to observed differences in BMP estimates.

Table 18. Sources of variability in BMP results, as standard deviation (NmL/g VS) based on a linear mixed-effects model.

Error source	Cellulose	Substrate A	Substrate B	Substrate C
Laboratory	28.8	22.6	27.8	39.6
Test	24.4	18.8	19.6	51.2

Notes: Residual standard deviation was 23.6 NmL/g VS.

Biases among laboratories (“Lab” row in Table 18) may be due to a myriad of factors, including biases inherent in measurement methods and equipment, performance of inoculum, and methods used for data processing. BMP values for this analysis were calculated using fixed VS contents for each substrate, so error in VS determination did not contribute here. However, biases between tests (“Test” row in Table 18) presumably do not include many of these sources, and it is significant that the model estimates of standard deviation are only slightly smaller for this source of error. Inoculum effects may have made a contribution to these biases, and may also explain differences among substrates.

4.9 Factors affecting BMP

Measurement conditions (Table 19, all factors) and substrate VS mass all have the potential to affect BMP (as do other, unmeasured, variables). However, levels were not randomly assigned, and it is likely that they were correlated with unmeasured differences between laboratories, and so it is not possible to conclusively show whether or not BMP measurements were affected by any particular factor. Additionally, with the lack of balance and relatively large number of possible predictors, various models could explain the observed results equally well. We used the following approach to attempt to extract information from this data set. Because mixed-effects models showed strong evidence of differences in variability among substrates, we analyzed each substrate separately. Separate analyses were also carried out for each test, to avoid pseudo-replication but still use least-squares multiple regression. The best-replicated level for each factor was taken as the reference level: wastewater treatment plant (WWTP) CSTR for inoculum source, mesophilic temperature for > 2 d for inoculum storage, and AMPTS for measurement method. Headspace flushing gas was not included in the analysis due to several missing observations. Substrate VS mass was included as a binary variable: either > 2 g as required in the protocol or not. Similarly, ISR was included as a binary variable: 1.5 < ISR < 6 or not. Mixing was confounded with measurement method - those labs using AMPTS or similar equipment all used continuous mechanical mixing, while all others used manual mixing - and was therefore was not included in any models. However, mixing may have contributed to observed differences between measurement methods.

All possible predictors (Table 19) were considered for each data subset, and a single “best” model was identified using a stepwise elimination procedure based on Akaike’s Information Criterion (AIC). Results described below are based on these resulting models. Because of the limitations of this data set, models cannot conclusively show the presence or magnitude of effects, but can provide evidence of effects. Results showed that all subsets had some correlation between predictors and measured BMP (Table 19). Measurement method was included in 6 of 8 models. The most consistent effects were lower BMP when manometric or non-AMPTS volumetric methods were used. The estimated effects ranged from 12 to 68 NmL/g VS below AMPTS results.

Inoculum source was retained in all but 2 models, which suggests that it may be important. A laboratory batch source of inoculum appeared to have the single largest effect on measured BMP in 2 of 8 models, but replication was low (2 labs for test 1, 1 for test 2). Other sources did not show consistent apparent effects. Similarly, inoculum storage was included in 5 models, but none of the effects had the same signs for all models. Other apparent effects were not consistent.

Table 19. Linear model coefficients for BMP (NmL/g VS) with predictor selection based on AIC [1]

Substrate	Cellulose		Substrate A		Substrate B		Substrate C	
Test	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
Intercept	387	350	385	379	391	361	464	512
Measurement method [2]								
Gas counter	29.3		17.2	-11.6	14		75.8	112
Other volumetric	**-17.3**		**-67.4**	**-6.73**	**-28.4**		**-30.1**	**-18**
Manometric	**-62.1**		**-68.7**	**-43.9**	**-47.4**		**-11.5**	**-31.7**
Gas chromatography	-25.2		25	0.9	42.2		88.9	38.4
Infrared	38		34.4	-8.2	14.8		-99.3	-124
Gravimetric	-40.1		-4.4	-21.6			23	-12.2
Inoculum source								
CSTR manure	-7.5	0.1	-8.1		-14.5	-3.1	-6.6	
UASB	-24	168	-77		-0.6	41.9	-147	
Lab batch	-78.5	-53.8	-48		-51.2	-102	-84.9	
CSTR codigestion	-13.2	37.4	-19		-11.4	32.4	28.3	
PF thermo.	-29.5	180	-4.0		0.5	61.7	-39	
Inoculum storage								
< 2 d	-32.8	29.3	9.2	26.9		58.7		
Ambient > 2 d	-20.2	33.7	-6.9	-4.5		6.9		
Thermo. > 2 d	41.8	254	-16.8	94.9		62		
4°C	-6.7	24.4	-12.1	44.5		50.2		
ISR < 1.5 or ISR > 6	22.1	-153	51.5	50.1			79.4	
Substrate VS < 2 g		-39.6	55.5					

[1] Models were fitted separately to each substrate/test combination (each column contains all the coefficients for a single model). Because all predictors are factors that were entered as dummy variables, the intercept term is the mean BMP for the reference conditions. Missing coefficients indicate that addition of the predictor did not improve AIC. Terms with a consistent effect for all subsets and replication of at least 5 labs are shown in bold.
[2] See Table 20 for the number of laboratories that used each method.

4.10 Factors that affect test validation

The discussion above (Section 4.9) on the challenges of variables selection apply to analysis of validation as well as BMP values, and a similar approach was used for identification of predictors that may affect the probability of validation. In this case, the response variable was the binary variable validation (whether or not a result was validated, based on the criteria described in section 3) and logistic regression was used. Here, results from different substrates are not completely independent since validation of any substrate requires validation of the cellulose results.

The fraction of validated observations varied substantially among factor levels (Table 20). For example, for substrate C in test 1, 8 out of 13 results (61%) from AMPTS II measurements were validated, while only 1 out of 6 (17%) made by other volumetric methods was validated. Differences this large and the correlation between measurement method and measured BMP described in Section 4 suggest effects of measurement methods and possibly other predictors on the probability of validation.

Table 20. Observed counts of observations (n = 3 batches) meeting all original validation criteria, grouped by test, measurement type, and substrate.

Measurement method	Test	Number of labs	Number of observations validated[1]			
			Cellulose	Substrate A	Substrate B	Substrate C
AMPTS II	1	13	8	7	7	8
AMPTS II	2	11	6	5	5	6
Gas counter	1	2	0	0	0	0
Gas counter	2	2	0	0	0	0
Other volumetric	1	6	1	0	0	1
Other volumetric	2	5	1	1	1	1
Manometric	1	6	1	1	1	1
Manometric	2	7	1	1	1	1
GC	1	2	0	0	0	0
GC	2	2	0	0	0	0
IR	1	2	0	0	0	0
IR	2	2	0	0	0	0
Gravimetric	1	1	0	0	0	0
Gravimetric	2	1	1	1	1	1
Unknown	1	1	0	0	0	0
Unknown	2	1	0	0	0	0

[1] Note: Each laboratory provided one observation (or in some cases none) per substrate.

But samples sizes are small for some factor levels, and so it is difficult to confirm that effects were present. For this particular measurement method comparison, for example, the probability of a difference this large when population proportions are actually equal is 0.18 (based on two-sample proportion equality test with Yates' continuity correction).

Overall validation of BMP results based on the original criteria (Section 4.3) was not clearly related to any predictors in test 2, but in test 1 the probability of validation appeared to be related to both measurement method (for cellulose, substrate A, and substrate B) and inoculum source (for cellulose and substrate A) in some cases (results not shown). Most measurement methods showed a consistent negative effect on the probability of validation, compared to AMPTS. This result is presumably due to a lower probability of meeting the 85%-100% cellulose BMP criterion, which was lower for nearly all other measurement methods in most cases (but not for cellulose in test 2 and substrate B in test 1). Inoculum from sources other than WWTP CSTRs was correlated with a lower probability of meeting the cellulose criterion for both tests, and overall validation for only cellulose and substrate A in test 1. There was some evidence

that validation was related to storage conditions. The probability of meeting criterion 1 was generally not consistently related to any predictors.

5 General discussion

A first observation is the difficulty of obtaining all the results in the form requested. The laboratories used to carry out the tests according to their own methods did often not follow the commonly agreed on protocol, e.g. TS and VS not in triplicate or no verification of inoculum quality according the different parameters defined for this purpose. The same was true for the data reporting which made it difficult to harmonize the data files which was very important for data analysis.

The present study was rather similar to the German study carried out by VDLUFA, except that the goals were different. Whereas the inter-laboratory study of VDLUFA has rather the purpose to validate the competence of the testing laboratories on the basis of a standardized method, a method was sought here to make the test result reliable for an analysis in any laboratory.

Although RSD for intra- and inter-laboratory reproducibility provide an indication whether in general the results obtained among the participating laboratories are overall of good quality, it is more important to consider the range of results obtained in the study. In reality, the analysis of the BMP is assigned to a single laboratory, and the question for the user is: "what is the maximum error that I can expect if I have a sample analyzed at any laboratory?".

The examination of the statistical parameters of our study, before taking into account the validation criteria, shows a significant dispersion. Z-scores, which represent a standardized estimate of laboratory bias, are found to be unfavorable for about one-third of the laboratories.

An important point for the examination of the dispersion parameters is the treatment of outliers. In this study the part was taken to eliminate only measures that do not involve comparison with other laboratories. Hence, we did not remove any outlier on too large a gap to the robust mean. Indeed it seemed important to keep only criteria that can be applied to a measurement in a single laboratory

The application of the validation criteria considerably reduces the range of the measurements. The original criteria for the study showed that only one quarter of the tests were validated, suggesting that they may have been too restrictive. A revision of these criteria was therefore considered and applied. The revised criteria allow for a smaller reduction in the range of the measurements, but they allow for the validation of approximately 60% of the tests, which is twice as much as with the original criteria.

However, the resulting range of values remains too large for reliable use of BMP results. It appears that the definition of these criteria should be further optimized and the test protocol defined strictly not leaving too much freedom to the testing laboratory.

There was evidence that both measurement method and inoculum substantially affected BMP values, and both factors could have contributed to observed inter-laboratory variability. Observed test biases may be related to inoculum, since other factors were presumably unchanged between tests carried out at the same laboratory. The studies conducted to date on differences among inocula are contradictory and do not confirm or refute an effect. It has to be noted that some laboratories reported very different results between the two sets of tests, a priori performed with the same type of inoculum. This suggests that other sources of error are involved, or that inoculum quality from individual sources varies substantially over time.

Differences between AMPTS and manometric and other volumetric results deserve some attention. Determining whether differences truly exist, and whether they are due to differences in methane production resulting from, e.g. differences in mixing (continuous mechanical versus intermittent manual), or error in determination of methane production, could help to address differences among laboratories. Designed experiments comparing multiple methods in a single laboratory, where other sources of error are minimized, could shed light on this problem.

It should also be remembered here that the principle of measurement is based on a number of assumptions:

– *The inoculum contains all the trophic groups necessary for the complete degradation of the substrate, even if some groups are initially little (or not) active:*

  In this case the origin of the inoculum does not affect the final result but only the kinetics of degradation, and laboratories with different inocula can carry out supposedly similar tests.

– *The production of methane is strictly additive:*

  The production of the substrate is obtained by subtracting the production of the inoculum alone from the production of (inoculum + substrate). However, it is possible that the addition of the substrate, generally at least a part of which is rapidly degradable, modifies the degradation of the residual organic compounds present in the inoculum (for example the delay at the moment when there is no longer any readily available organic matter). The withdrawal of the production of the inoculum alone would then not correspond to the reality.

– *The ISR expressed in VS is representative and comparable for the different inocula and substrates:*

  This is not really a hypothesis as it is obvious that the VS content of the inoculum does not represent the active microbial biomass. The microorganism concentration of the inoculum is not known because it is difficult to measure. We know that this is only a very rough estimate since a large part is represented by recalcitrant organic matter of the original substrate. To date there is no simple measure to express this ratio with a better representativeness.

– *There is no nutritional limitation or inhibition during the test and material transfers are not limiting*.

6 Concluding remarks

This international study on BMP tests has shown that the application of validation criteria as defined in the guidelines published in WST & T considerably decreased the range of BMP test results. On the other hand, the range remains significant despite the application of a standard set of conditions and validation criteria.

The obtained results would therefore not make it possible to use the BMP thus measured with the initial objectives to predict the production on an industrial installation from a set of substrates, or to use the BMP as a parameter to validate the proper functioning of an installation. Indeed the variability in BMP measurements is still much too large.

We have highlighted in this study the aspects on which optimization is possible:

– further standardize the BMP test protocol

– define more suitable validation criteria

– discuss with the laboratories presenting the most important deviations from the average in order to identify their origin

One point remains, the influence of the inoculum used for the test on the results, since each laboratory uses a different inoculum and it is only very little characterized.

In order to continue the standardization and finally make the BMP tests reliable, it was decided that a second workshop should be organized and a second inter-laboratory study carried out. The workshop will take place in April 2018 in Freising, Germany. A second inter-laboratory study could then be envisaged for Fall 2018.