

Article

Improving Inter-Laboratory Reproducibility in Measurement of Biochemical Methane Potential (BMP)

Sasha D. Hafner ^{1,*},[†] , H el ene Fruteau de Laclos ², Konrad Koch ³  and Christof Holliger ^{4,*}

¹ Department of Engineering, Aarhus University, Finlandsgade 12, 8200 Aarhus N, Denmark

² Methaconsult, 1028 Pr everenges, Switzerland; helene.fruteau@methaconsult.com

³ Chair of Urban Water Systems Engineering, Technical University of Munich, Am Coulombwall 3, 85748 Garching, Germany; k.koch@tum.de

⁴ Ecole Polytechnique F ed erale de Lausanne EPFL, School of Architecture, Civil and Environmental Engineering, Laboratory for Environmental Biotechnology, Station 6, 1015 Lausanne, Switzerland

* Correspondence: sasha@hafnerconsulting.com (S.D.H.); christof.holliger@epfl.ch (C.H.)

† Present address: Hafner Consulting LLC, Reston, VA 20191, USA.

Received: 14 May 2020; Accepted: 16 June 2020; Published: 19 June 2020



Abstract: Biochemical methane potential (BMP) tests used to determine the ultimate methane yield of organic substrates are not sufficiently standardized to ensure reproducibility among laboratories. In this contribution, a standardized BMP protocol was tested in a large inter-laboratory project, and results were used to quantify sources of variability and to refine validation criteria designed to improve BMP reproducibility. Three sets of BMP tests were carried out by more than thirty laboratories from fourteen countries, using multiple measurement methods, resulting in more than 400 BMP values. Four complex but homogenous substrates were tested, and additionally, microcrystalline cellulose was used as a positive control. Inter-laboratory variability in reported BMP values was moderate. Relative standard deviation among laboratories (RSD_R) was 7.5 to 24%, but relative range (RR) was 31 to 130%. Systematic biases were associated with both laboratories and tests within laboratories. Substrate volatile solids (VS) measurement and inoculum origin did not make major contributions to variability, but errors in data processing or data entry were important. There was evidence of negative biases in manual manometric and manual volumetric measurement methods. Still, much of the observed variation in BMP values was not clearly related to any of these factors and is probably the result of particular practices that vary among laboratories or even technicians. Based on analysis of calculated BMP values, a set of recommendations was developed, considering measurement, data processing, validation, and reporting. Recommended validation criteria are: (i) test duration at least 1% net 3 d, (ii) relative standard deviation for cellulose BMP not higher than 6%, and (iii) mean cellulose BMP between 340 and 395 $NmL_{CH_4} g_{VS}^{-1}$. Evidence from this large dataset shows that following the recommendations—in particular, application of validation criteria—can substantially improve reproducibility, with $RSD_R < 8\%$ and $RR < 25\%$ for all substrates. The cellulose BMP criterion was particularly important. Results show that it is possible to measure very similar BMP values with different measurement methods, but to meet the recommended validation criteria, some laboratories must make changes to their BMP methods. To help improve the practice of BMP measurement, a new website with detailed, up-to-date guidance on BMP measurement and data processing was established.

Keywords: biomethane potential; BMP; laboratory methods; interlaboratory test; intra-laboratory reproducibility; round robin test; methane yield; method standardization; validation criteria; measurement repeatability; BMP dataset

1. Introduction

Determination of the biochemical methane potential (BMP) of organic substrates is a key component of anaerobic digestion (AD) research as well as application of AD at full scale. In laboratory research, BMP tests are commonly used to investigate the effect of pre-treatment on methane potential or production rate [1,2]. BMP measurements have been shown to be useful for evaluation of full-scale performance [3,4], and BMP is used to assess substrate quality and predict CH₄ production by full-scale biogas plants. Additionally, residual BMP of digestate is used to estimate CH₄ emission to the atmosphere [5,6].

The original protocol for measurement of BMP was published 40 years ago [7]. Since then, many variations of the method have been described, with different techniques for measuring the volume of biogas produced and its composition [8–13]. Although originally described as a “relatively simple bioassay” [7], inter-laboratory studies have shown that BMP of the same substrates measured by different laboratories varies widely, suggesting that test protocols are not sufficiently standardized [14–16]. The implication from these studies is that there is high uncertainty in any single BMP value obtained from one laboratory. Given that research, business, and regulatory decisions are based on these measurements, there is a general consensus among biogas professionals that accuracy must be improved.

In the early 2000s, a task group of the AD Specialist Group of the International Water Association (IWA) and the Association of German Engineers (VDI) both undertook efforts to harmonize BMP and, more generally, anaerobic biodegradability tests. The former group consisted of international AD experts, and it ultimately published review articles and proposed a protocol for measuring BMP of solid organic wastes and energy crops in 2009 [17], whereas the VDI published a standard in 2006 [18], with an updated version in 2016 [11].

Because of continuing problems with BMP reproducibility [14,15], a third initiative was launched in June 2015 with an international workshop in Leysin, Switzerland, that included participants from 31 laboratories. Based on roundtable discussions during this workshop and intensive email exchange among all participants, a new set of guidelines for measurement of BMP was developed [19]. Similar to the VDI guidelines, actions and criteria considered to be compulsory in order to validate a BMP test result were presented along with recommendations. According to the resulting Leysin guidelines [19], validation of BMP results requires: triplicate bottles for blanks (with only inoculum) and all substrate bottles (with inoculum and substrate), inclusion of a positive control in addition to the blank assays, termination of tests only when daily methane production rate during three consecutive days is <1% of the accumulated volume of methane, that the BMP of the positive control is between 85% and 100% of the theoretical maximum BMP, that relative standard deviation (RSD) in methane (CH₄) production by blanks is <5%, and that the RSD among the triplicates for mean BMP (RSD_B) is <5% for the positive control and homogenous substrates or <10% for heterogeneous substrates. Recommendations were intended to help increase the likelihood that test results can be validated and address the inoculum source and properties, substrate preparation and characterization, test setup, and data analysis and reporting.

Although there was consensus among the Leysin workshop participants on the compulsory actions and criteria for BMP test validation, there was much less agreement about the primary sources of observed variability in BMP test results obtained in inter-laboratory studies. Recent reviews have summarized the work carried out to study the influence of key factors such as inoculum, substrate, experimental and operational conditions, and data analysis and reporting [20–22]. Specific work addressed, for example, inoculum origin [23–26], inoculum carryover and dilution [27], and inoculum storage [28–30] effects. Results from these studies show convincingly that the selection and storage of an inoculum influences the rate of methane production, but effects on the ultimate methane potential, i.e., BMP, have varied from negligible to large. The inoculum-to-substrate ratio (ISR) has also been proposed to have an influence on BMP, but here as well, large effects have been observed in some studies [26,31] and not others [32,33]. Other studies have focused on differences among measurement techniques [25,34–38]. In particular, manometric BMP measurements have been found to be sensitive

to headspace volume and pressure [34–36]. Surprisingly large differences (up to 2-fold) have been found between different volumetric methods as well [37]. All BMP methods except the gravimetric approach [12,39] could be affected by leakage [36,40], which is rarely quantified. Data processing, i.e., calculation of BMP from raw measurements, may be another source of variation. This may include differences in equations, e.g., whether gas volume standardization includes water vapor [11,39] or not [10,41], but also whether local conditions are used in standardization [13]. Reducing error in BMP measurement will require a better understanding of these effects. Unfortunately, assessment of some effects is difficult to separate from measurement errors. A summary of the literature might conclude that anything can affect BMP, but it is probable that literature results reflect both publication bias [42] as well as errors by inexperienced researchers.

The present work continues the task of improving the quality of BMP measurement, and builds on previous work—in particular the results and network from the Leysin workshop and the associated guidelines [19]. This paper presents results and interpretation from two related international inter-laboratory studies on BMP (IIS-BMP), collectively referred to as the “IIS-BMP project”. In total, 37 laboratories, primarily from Europe, participated. Project objectives were as follows:

1. Quantify and partition observed variability in BMP measured using a standardized protocol;
2. Assess possible sources of error in BMP measurement, including inocula, calculation errors, and systematic measurement biases;
3. Test and revise BMP validation criteria based on a quantitative analysis of collected data.

2. Materials and Methods

2.1. Project Structure

Two international inter-laboratory studies on BMP were carried out. In the first study (S1), a BMP protocol based on Holliger et al. [19] was tested by measuring the BMP of three homogenous substrates and microcrystalline cellulose in two tests (T1 and T2). Results were discussed at a workshop in Freising, Germany, in 2018, and subsequently a second study (S2) with a single test was carried out by each participating laboratory. S2 BMP tests followed a similar protocol but also included efforts to assess both inoculum and measurement method effects on BMP. Detailed data were submitted for both S1 and S2, including not only BMP values calculated by each participating lab (referred to as “reported” BMP values) but also all raw laboratory data, allowing for calculation of specific methane production (SMP) curves and therefore BMP at any duration by the project organizers.

2.2. Participating Laboratories

Both research-oriented and commercial laboratories participated in this project, although most were primarily research-oriented. The network of participating laboratories was initially based on the group that participated in the Leysin workshop described in Section 1 [19]. New participants joined for S1, and some joined or left for S2, but most laboratories (31) participated in both. In total, 37 laboratories from the following 14 countries participated: Australia, Belgium, Canada, Czech Republic, Denmark, France, Germany, Italy, Netherlands, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.

2.3. Test Substrates

Four dry, finely ground, and homogeneous test substrates were selected. All substrates were analyzed by Eurofins Scientific AG (Schönenwerd, Switzerland) for elemental composition (following ISO 16948 for C, H, and N [43], and ISO 16993 for O [44]). Elemental composition was used to estimate maximum theoretical BMP (Table 1) from stoichiometry using the `predBg()` function from the `biogas` package (version 1.24.3) [45,46] in R (version 3.6.3) [47]. The calculation was based on the approach described by Rittmann and McCarty [48], Equation (13.5), assuming 100% degradation and no use of substrate for microbial biomass production. The substrates A, B, and C (referred to

as SA, SB, and SC here) used in S1 were commercial animal feeds from Cargill Feed & Nutrition (Geneva, Switzerland). Substrate A was a pig feed that contained wheat, triticale, barley, peas, bran, rapeseed, soya, fat, and amino acids. Substrate B contained cellulose, starch and nitrogen-containing compounds. Substrate C was a fodder called “Probos” that contained cod liver oil, sunflower seed, rapeseed, soya, flaxseed, wheat germ, wheat bran, wheat flour, oat, and yeast. In addition to SC, a less degradable substrate was included in S2: substrate D (SD) consisted of only ground wheat straw. Microcrystalline cellulose (referred to as CEL here) (CAS Number 9004-34-6, Sigma-Aldrich Chemie GmbH, Buchs, Switzerland), also provided by the organizers, was used as a positive control in both S1 and S2. Maximum theoretical BMP was between 400 and 500 $\text{NmL}_{\text{CH}_4} \text{g}_{\text{VS}}^{-1}$ for all substrates except lipid-rich SC (Table 1). Participating laboratories had no knowledge of the source or composition of the substrates apart from CEL.

Table 1. Substrates used for biochemical methane potential (BMP) tests.

Substrate Key	Tests	Description	TS (% FM) *	VS (% TS) †	Chemical Formula ‡	Theoretical Max. BMP ($\text{NmL}_{\text{CH}_4} \text{g}_{\text{VS}}^{-1}$) ¶
SA	T1	Animal feed	88.8	93.3	$\text{C}_{17}\text{H}_{30}\text{O}_{12}\text{N}$	440
SB	T1	Animal feed	89.1	97.2	$\text{C}_{23}\text{H}_{37}\text{O}_{16}\text{N}$	448
SC	T1, T2	Animal feed	92.8	87.9	$\text{C}_{18}\text{H}_{31}\text{O}_8\text{N}$	606
SD	T2	Wheat straw	92.2	93.8	$\text{C}_{59}\text{H}_{88}\text{O}_{38}\text{N}$	481
CEL	T1, T2	Microcrystalline cellulose	94.9/99.0	100.0	$\text{C}_6\text{H}_{10}\text{O}_5$	414

* Total solids as percentage of fresh mass (median of values measured by participating laboratories). † Volatile solids as a percentage of TS (median of values measured by participating laboratories). ‡ Chemical formula for cellulose, otherwise empirical chemical formula. ¶ Theoretical maximum biochemical methane potential based on elemental composition (Section 2.3). || TS was higher in S2.

2.4. BMP Measurement

2.4.1. First Study (S1)

Thirty-three laboratories participated in S1. Each laboratory determined the BMP of three substrates following a protocol based on Holliger et al. [19]. Two tests had to be carried out about one month apart in order to quantify both inter- and intra-laboratory reproducibility.

Substrate TS and VS were determined in triplicate by each laboratory by drying at 105 °C and igniting at 550 °C [49,50]. Each BMP test was carried out in triplicate with blanks (inoculum only) and CEL as a positive control. There were no restrictions on the origin of the inoculum, but the protocol stated that it had to contain a “highly diverse” methanogenic community. Inoculum VS content should have been between 15 and 40 $\text{g}\cdot\text{L}^{-1}$, the pH between 7.0 and 8.5, total volatile fatty acids < 1.0 $\text{g}\cdot\text{L}^{-1}$ (as acetic acid), total ammonia-nitrogen concentration < 2.5 $\text{g}\cdot\text{L}^{-1}$, and alkalinity > 3 $\text{g}\cdot\text{L}^{-1}$ (as CaCO_3). Sieving was not permitted, and storage (at ambient or test temperature) had to be ≤ 5 days.

Each BMP bottles had to contain at least 2 g of substrate, along with trace element and vitamin solutions (according to the recommendations by Angelidaki et al. [17]), which were provided by the project organizers. Total VS concentration was to be between 20 and 60 $\text{g}\cdot\text{L}^{-1}$. Headspace was to be flushed with a mixture of N_2 and CO_2 (20–40% CO_2) or N_2 (with a check of pH change in one bottle). Three substrates (SA, SB, and CEL) were tested with a VS-based inoculum-substrate ratio (ISR) of 2:1, the third, SC, with an ISR of 4:1. The incubation temperature was 35 ± 2 °C, and the bottles were mixed either manually (at least once per day requested, in practice weekends were likely omitted) or automatically. There was no restriction on the way methane production was determined; however, ambient temperature and pressure had to be recorded at each measurement time in order to be able to standardize the gas volume (to dry volume at 0 °C and 101.325 kPa). Tests were run at least until daily net (i.e., with inoculum contribution subtracted) CH_4 production dropped below 1% of the cumulative value for at least 3 consecutive days [19] (Section 2.6.1).

2.4.2. Second Study (S2)

Thirty-seven laboratories participated in S2. The only required substrates were SC and CEL, while testing SD was optional. SC was chosen because preliminary results suggested it had low reproducibility in S1, potentially because of its high lipid content. SD was rich in lignocellulosic materials and therefore slowly degradable, in contrast with all the other substrates.

The S2 protocol was similar to the one of S1, but with a few differences, described in the following. As for S1, there were no restrictions on the origin of the inoculum and the quality criteria were the same. However, pre-treatment such as sieving was allowed if needed. A smaller mass of substrate (1 g VS) was permitted in S2, with an ISR of 2:1 for SC and 1:1 for SD. No restriction was given for total VS concentration (although in most cases values were within the 20 to 60 g·L⁻¹ range from S1). The incubation had to be done at a mesophilic temperature (35 to 40 °C, with a maximum of 2 °C variation) that matched the temperature of the digester from which the inoculum was taken.

To test the influence of the inoculum on BMP, the majority of laboratories tested at least two inocula. Expecting delays due to customs control when sharing inoculum across country borders, inoculum exchange was done at the national level only. In three countries, one laboratory from each sent an inoculum from a single source (referred to as “shared” in results) to all other laboratories within the same country and all carried out the BMP test during the same time. Resulting BMP values were compared to those from each laboratory’s own unique routine inoculum (referred to as “own” in results). In three other countries, an exchange of routine inocula was organized between pairs of laboratories (two per country). Additionally, three laboratories compared two or more measurement methods.

2.4.3. Measurement Methods

BMP measurement methods used by participating laboratories were grouped into five categories: an automated volumetric method using the system produced by Bioprocess Control (Lund, Sweden) called the Automatic Methane Potential Test System II (AMPTS II, 15 laboratories) [13,51], other volumetric methods (14 laboratories) [11,52], manometric methods (10 laboratories) [11,52], the absolute gas chromatography (GC) method (2 laboratories) [8], and the gravimetric method (2 laboratories) [12,39]. Most non-AMPTS II methods were manual (7 were automated), generally with biogas accumulation under pressure in the headspace of BMP bottles during incubation intervals. Laboratories using manual volumetric or manometric methods were asked in S2 to include measurement of initial and final bottle mass, providing data for a gravimetric evaluation some results [12]. Manometric, gravimetric, and most non-AMPTS II volumetric measurements included measurement of CH₄ concentration along with biogas quantity (by pressure, mass loss, or volume, respectively) for each sampling event. The concentration of CH₄ within biogas was determined using gas chromatographs or infrared analyzers, but method details were not collected. One laboratory in S2 used a new gas density method [9]. In contrast, AMPTS II units and some other systems remove biogas CO₂ using an alkaline trap before measuring gas production, and provide cumulative standardized CH₄ volume directly by correcting for measured temperature and pressure. Absolute GC measurements are proportional to the total quantity of CH₄ in each bottle, and neither CH₄ concentration nor total biogas quantity is determined [8].

2.5. Data Submission

The primary unit of observation for data collection was a measurement interval for a single bottle, but submitted data included variables from four levels: test, substrate, bottle, and measurement interval. Test level variables included at least TS, VS, pH, and origin of inoculum; biogas measurement method; headspace flushing gas (e.g., 100% N₂ or 80:20 N₂:CO₂); and incubation temperature. Substrate level variables were substrate TS and VS at the replicate level, along with BMP and associated relative standard deviation RSD_B (which was to include contributions from both blanks and bottles with substrate [53]). Inoculum and substrate mass (fresh basis) and, for some methods, headspace volume

were reported for each individual bottle. Lastly, measurement interval data contained the variables necessary for determining CH₄ production by each individual bottle within each measurement interval. Participants entered data into spreadsheet templates provided by the project organizers (templates are available in the Supplementary Material).

2.6. Data Analysis and Calculations

2.6.1. Data Processing

Cumulative CH₄ production was calculated from laboratory measurement for all measurement intervals using the biogas package (version 1.24.3) [45,46] in R (version 3.6.3) [47]. Volumetric data (including manual and automated approaches) were processed following the standard approach [54] using the calcBgVol() function. Calculations for manometric measurements followed the standard approach [55] and were done with the calcBgMan() function, assuming saturation with water vapor at all times. Volumetric and manometric data included two types of measurements: those with CH₄ concentration normalized to CH₄ and CO₂ only and others with as-measured values (ostensibly corrected only for moisture), which require separate estimation of vented (removed) and residual headspace CH₄. These measurements were processed as described using methods 1 and 2 [54,55], respectively (cmethod = “removed” and cmethod = “total” in the functions, respectively [45,46]). Method 2 requires the headspace volume of each bottle, which was provided by participants. Gravimetric data were processed using the calcBgGrav() function, which implements the calculations described previously [12,56], with a correction included for the initial headspace mass. For the gravimetric evaluation of a subset of manometric and volumetric results (subset H, Section 2.6.2), the vol2mass() function was used to calculate biogas density and expected bottle mass loss from volume- or pressure-based results. A correction was included for the initial headspace mass, with density calculated based on flushing gas composition using the gasDens() function. The cumBg() function was used for absolute GC measurements, with calculations following Hansen et al. [8]. Gas density measurements were processed using the calcBgGD() function, which implements the calculations described in Hafner et al. [57], using “GD_t” values for mass loss and gas volume as described in Justesen et al. [9], Section 2.1.3.

The primary unit of observation for data analysis was a BMP value for a particular substrate from a single test carried out by a single lab. Each BMP value was calculated from measurements made on 3 bottles with substrate and inoculum and 3 bottles with inoculum only (blanks). BMP was calculated using data on bottle contents (mass of inoculum, identity, and mass of substrate) along with cumulative CH₄ production using the summBg() function, which follows the details given in Hafner et al. [53]. Several BMP durations were used for each substrate within each individual test including both fixed and relative durations. Four relative durations were considered, each defined by a response variable, a maximum relative rate, and a minimum measurement interval over which the maximum rate cannot be exceeded. The term “1% net 3 d duration” is used for the incubation duration when the daily net CH₄ production (or average interval rate in NmL_{CH₄} d⁻¹ for those manual methods with sampling frequency < 1 d⁻¹) dropped below 1% of cumulative net production for a continuous period of at least 3 consecutive days. Net CH₄ production was calculated by subtracting the inoculum contribution from gross (total) CH₄ production, as in calculation of SMP and BMP [53]. Other durations that were considered were: 1% gross 3 d, 0.5% gross 3 d, and 0.5% net 3 d. The 1% net 3 d duration was recommended by Holliger et al. [19], and the more conservative 0.5% gross 3 d approach was recommended by VDI [11]. Durations were specified in the summBg() function using the “when” and “rate.crit” arguments. RSD_B was calculated including contributions from three sources: substrate bottles, blanks, and substrate VS measurement [53].

2.6.2. Data Subsets

Ten related data subsets were used for data analysis, with each serving a different purpose. All subsets contained data from multiple laboratories, but not all laboratories were represented in each subset. Reported BMP values (subsets A or B) are typically the only results available in inter-laboratory comparisons. All subsets, including measured or calculated BMP and associated variables, are available as part of the supplementary data, with the exception of G, which was excluded because of concerns about confidentiality when a country key, inoculum key, and measurement method are provided. The 10 subsets are described below.

- A. Reported BMP. For S1, submitted standardized CH₄ volume at the duration identified by participants (1 value per bottle) was used to calculate BMP along with data on bottle contents and laboratory-measured VS concentrations. For S2, participating laboratories provided BMP directly in spreadsheet templates. For both studies, participants identified the time that met the requested duration criterion based on their own calculations.
- B. Reported BMP under reference conditions. Thus, a subset of A. For 5 laboratories that used multiple biogas measurement methods, only results from their typical method were retained (with the exception of one laboratory that used both manometric and gravimetric methods, because of a dearth of gravimetric results). ISR was limited to values given in the study protocol (Sections 2.4.1 and 2.4.2), and mean BMP values based on less than 3 substrate bottles or less than 3 blanks were excluded. However, the use of multiple inocula per laboratory was permitted, and therefore, there was an unequal number of BMP values among laboratories. Because of apparent difficulty in exactly matching requested ISR values, limits were somewhat lenient: nominal ISR was taken as 2 for 1.5 < ISR < 2.5, and 4 for 3 < ISR < 5. For CEL, ISR > 0.7 was accepted.
- C. Calculated BMP at reported durations. Mean BMP values ($n = 3$ substrate bottles) were calculated from submitted raw data at durations that exactly matched those in set A, using laboratory-measured VS concentrations. These values were aligned (merged) with reported values to assess calculation and data entry errors by participants. Values that could not be matched and any with missing replicates were dropped.
- D. Calculated BMP at various durations. BMP values were calculated from all submitted raw data and median substrate VS concentrations for various relative and fixed durations. This subset was used to compare duration criteria.
- E. Calculated BMP under reference conditions. A subset of D, but with the same constraints as in subset B. Only 1% net 3 d (E1) and final (latest available) durations (E2) were used.
- F. Calculated BMP with laboratory-measured VS. As with E1 but calculated with laboratory-measured substrate VS concentrations. This set was compared to E1 to assess the contribution of VS measurement error to inter-laboratory variability in BMP.
- G. Calculated BMP from inoculum comparisons. A subset of D from trials where inocula were compared. BMP was taken as 1% net values. This set was divided into two for analysis based on how the inoculum exchange was carried out: a single inoculum shared among more than two labs in each country (G1) or exchange of two inocula between two labs per country (G2) (Section 2.4.2).
- H. Calculated CH₄ production with gravimetric evaluation. Results from tests where the initial and final mass of each bottle was determined by participating laboratories. Only the final (latest) values were used in order to evaluate measurements using total mass loss over the duration of the trial. Unlike the other subsets, the response variable of interest was total CH₄ production, calculated for each bottle.
- I. Calculated BMP from 3 laboratories which compared measurement methods. Unlike the other subsets, 1% net 3 d BMP was calculated separately for each bottle so methods could be compared. One laboratory compared three methods, while the others used two. Four methods were used in total: automated volumetric (AMPTS II and a different system) and manual volumetric, manometric, and gravimetric.

- J. Calculated BMP at the 1% net 3 d duration from 14 laboratories that varied ISR within BMP tests. A subset of D.

2.6.3. Data Analysis and Display

Inter-laboratory reproducibility was quantified as relative standard deviation (percentage of overall mean value) for a subset of BMP values for a single substrate and test (Section 2.6.2) and is represented by RSD_R . Relative range (the difference between the maximum and minimum values expressed as a percentage of the mean) was also used and is represented by RR. No attempt was made to detect or remove outliers prior to calculation of RSD_R or RR. However, for the purpose of summarizing results, “extreme values” were defined as those more than 25% above or below the median response.

Boxplots (box-and-whisker plots) were used for graphical comparisons. In all boxplots, the heavy line shows the median, the box shows 25th and 75th percentiles, and vertical lines (whiskers) show the range, excluding outliers, which are plotted as points. In these plots, outliers were identified as values beyond the box by more than 1.5 times the interquartile range. To facilitate comparisons but still show the presence of outliers, extreme values were adjusted to the following limits: CEL, 250–450; SA, 300–450; SB, 250–450; SC, 250–650; and SD, 150–350 $\text{NmL}_{\text{CH}_4} \text{gVS}^{-1}$. These values were selected to show outliers without excessively large axis limits and apply only to the plots.

Mixed-effects models were used to estimate the contribution of laboratories and tests to observed BMP variability. The laboratory error term provides an estimate of the magnitude of variation in systematic biases among laboratories, i.e., an indication of inter-laboratory reproducibility. Although this source of error is likely systematic for a single lab, when evaluating it among multiple labs, it is expected to behave as a random source of error [58]. Laboratory and test were therefore included as random effects, with test as a three-level factor (S1 T1, S1 T2, and S2 T1) nested in laboratory. This implies no expectation for consistent differences among the tests across all laboratories, and therefore, the test term provides an indication of average intra-laboratory reproducibility within individual labs. Substrate type was included as a fixed effect with four levels. The response variable was \log_{10} -transformed reported BMP from set B, and the unit of observation was a single mean BMP for a single substrate from one test carried out by a single laboratory. A logarithmic transformation was used based on the expectation that random error in BMP is better represented by a lognormal distribution than a normal distribution and more likely proportional to mean BMP than fixed [59]. The `lmer()` function from the `lme4` package in R was used for parameter estimation based on a restricted maximum likelihood algorithm [60,61].

Linear models were used to explore correlation among BMP and possible predictor variables [62]. For this, the `lm()`, `aov()`, `summary.lm()`, and `summary.aov()` functions were used in R [47]. The response variable was calculated BMP from set E1, and as with the mixed-effects models, it was \log_{10} -transformed. The unit of observation was the same as with mixed-effects models. The primary predictor of interest was measurement method, and substrate was included as a covariate. Because it was not practical to assign measurement methods to participating laboratories (and no attempt was made to do so), this analysis was effectively observational (not experimental), with a real possibility of confounding [63]. Data from S1 and S2 were analyzed separately. Analysis of variance (ANOVA) was used with the second set of inoculum comparisons (subset G2) to test for an overall effect of inoculum origin and compare contributions to observed variability. Country, substrate, and an interaction between country and inoculum origin (to test inoculum effects by country) were included as factors.

Laboratories that compared measurement methods provided an experimental dataset (subset I) that was analyzed using both ANOVA and mixed-effects models to test for differences among measurement methods and evaluate their contribution to observed variability. The response variable was \log_{10} -transformed reported BMP from set I, and the unit of observation was a single BMP value for a single bottle. As above, the `lm()`, `aov()`, `summary.lm()`, `summary.aov()` functions were used in

R [47] for ANOVA [62], and the lmer() function from the lme4 package in R was used for mixed-effects models, based on a restricted maximum likelihood algorithm [60,61].

For comparisons between data subsets (Section 2.6.2), the nonparametric Wilcoxon test [64] was used through the wilcox.test() function [47]. The same test was used to compare standard deviation among BMP measurements (calculated by country and substrate) from the inoculum comparisons included in subset G1. For evaluating apparent measurement bias based on a gravimetric check, a one-sample *t*-test on the relative difference was used [59] with the t.test() function [47]. Correlation in BMP to CEL BMP was quantified using Kendall's tau [59], which is based on association in ranks and is not sensitive to outliers or deviation from a linear response. Robust regression was applied with the rlm() function to calculate the slope of the response [65].

2.6.4. Validation Criteria

The original validation criteria proposed by Holliger et al. [19] and two candidate revisions were evaluated by applying them to results from data subset E2 (Section 2.6.2), using BMP based on the latest available duration. For all sets, a duration at least as long as the 1% net duration (Section 2.6.1), $n = 3$ substrate bottles and 3 blanks, and the use of CEL as a positive control were required criteria. The following three sets of criteria were evaluated:

O. Original [19]:

- i. Duration at least 1% net 3 d;
- ii. Blank SMP $RSD \leq 5\%$;
- iii. Cellulose $RSD_B \leq 5\%$;
- iv. Substrate $RSD_B \leq 5\%$;
- v. Cellulose BMP between 352 and 414 $NmL_{CH_4} gVS^{-1}$.

R1. Revision 1 (lenient):

- i. Duration at least 1% net 3 d;
- ii. Cellulose $RSD_B \leq 6\%$;
- iii. Cellulose BMP between 320 and 414 $NmL_{CH_4} gVS^{-1}$.

R2. Revision 2 (recommended):

- i. Duration at least 1% net 3 d;
- ii. Cellulose $RSD_B \leq 6\%$;
- iii. Cellulose BMP between 340 and 395 $NmL_{CH_4} gVS^{-1}$.

3. Results and discussion

3.1. Data Set Overview

3.1.1. Data Set Size

In total, 444 BMP values ($n = 3$ bottles with inoculum and substrate for each observation, along with 3 blanks per BMP test) were submitted by 37 laboratories from 15 countries (Table 2). After calculation of BMP and removal of tests with non-reference values (Section 2.6.2), the primary dataset (E) contained 344 unique BMP values. A subset of 21 labs from 6 countries provided results of inoculum comparisons, consisting of 116 calculated BMP values (sets G1 and G2). The size of each dataset is given in Table 2.

Table 2. Size of each data subset.

Subset	No. Labs	No. Countries	No. BMP Tests	No. Observations *
A	37	15	124	444
B	36	14	117	359
C †	37	15	122	410
D ‡	37	15	123	412
E ¶ and F	36	14	116	344
G1	15	3	30	84
G2	6	3	12	32
H	7	7	14	154
I	3	3	12	109
J	14	9	22	66

* Number of BMP values based on $n = 3$ bottles with substrate, except for set H, where value is number of total CH_4 production values, each from a single unique bottle, or for I, where value is the number of BMP values calculated separately for each bottle in order to compare measurement methods within laboratories. † After dropping observations with no match in A. ‡ Count depends on when BMP was evaluated (not all tests continued to most conservative duration criterion), and these values are for 1% net duration. ¶ This subset was used for evaluation of validation criteria, and for that, 16 observations from 7 tests where cellulose was not included as a substrate were dropped, and only 35 laboratories were included for evaluation of criteria.

3.1.2. BMP Reproducibility

Reported BMP values showed low to moderate inter-laboratory variability: RSD_R ranged from under 8% up to 24% depending on substrate and test (Table 3, Figure 1). Estimates of sources of random error from mixed-effects models were 3.6% for tests and 5.3% (standard deviation) for laboratories (intra- and inter-laboratory reproducibility, respectively) after exclusion of extreme values (12 observations were eliminated from subset B based on residuals) (Table S8). Apparent degradability was high for all substrates except SD (wheat straw), presumably due to its lignocellulosic composition. The overall mean CEL BMP was $346 \text{ NmL}_{\text{CH}_4} \text{ gVS}^{-1}$ (S1 T1 and S2) to $365 \text{ NmL}_{\text{CH}_4} \text{ gVS}^{-1}$ (S1 T2), implying that many values were below the validation limit proposed by Holliger et al. [19] of $352 \text{ NmL}_{\text{CH}_4} \text{ gVS}^{-1}$ (see Section 3.4.1).

In general, variability was comparable to results from other studies. In the inter-laboratory comparison presented by Raposo [15], RSD_R was 15% for cellulose and as high as 37% for a simple proteinaceous substrate (gelatin). Similarly, Cresson et al. [14] found RSD_R of 13 to 21% for BMP of homogenous substrates measured in several French laboratories. In a 2017 German test summarized by Weinrich et al. [16], RSD_R was 8% for cellulose, but as high as 26% for other substrates, which included fresh maize silage, oat bran, and animal feed. Similarly, repeated German tests carried out for 9 years showed a cellulose RSD_R between 7% and 13% (excluding nearly 20% for a single year), with similar values for maize silage [16].

In addition to RSD_R , the range of BMP values is important, and unusual (extreme, including both low and high) values were present for all substrates (Figure 1). The RR in reported BMP values was $\geq 30\%$ for all substrates and $>110\%$ for SB and SD (Table 3). All of the highest extreme values were above theoretical maximum yields calculated from elemental composition, confirming unambiguous measurement error (Figure 1). Clearly, the presence of such high inter-laboratory ranges and, from the perspective of a researcher or plant manager, the possibility of obtaining an inaccurate BMP result for a single sample, is a significant motivation for BMP standardization. Although it is easy to identify unusual values in this or other inter-laboratory datasets, it is impossible to do so when one has only a single value provided by a single laboratory (unless it is higher than a known theoretical maximum BMP). These extreme values also inflate RSD_R .

Based on these results, it is suggested that a successful BMP protocol (including validation criteria) is one that can almost always deliver $\text{RSD}_R < 7\%$ and $\text{RR} < 20\%$ for BMP of homogenous substrates. Both of these values are arbitrary but would represent a significant improvement compared to reported BMP values here (Table 3) and in other studies [14–16].

Table 3. Summary of reported BMP values (subset B).

Study	Test	Substrate	No. Labs	No. Obs.	No. Extreme *	Mean BMP (NmL _{CH₄} g _{VS} ⁻¹)	RSD _R (%) †	RR (%) ‡
S1	T1	CEL	23	23	0	346	9.0	32
S1	T1	SA	22	22	0	368	8.9	37
S1	T1	SB	21	21	1	353	24.3	114
S1	T1	SC	20	21	1	465	12.9	50
S1	T2	CEL	22	22	2	365	11.4	58
S1	T2	SA	22	22	0	374	7.5	31
S1	T2	SB	22	22	2	372	22.6	121
S1	T2	SC	20	22	0	500	9.5	36
S2	T1	CEL	36	72	0	346	7.7	41
S2	T1	SC	35	65	3	494	13.1	81
S2	T1	SD	29	47	3	279	17.3	128

* Number of extreme observations (>25% difference from median). † Relative standard deviation (% of mean).

‡ Relative range (maximum—minimum as % of mean).

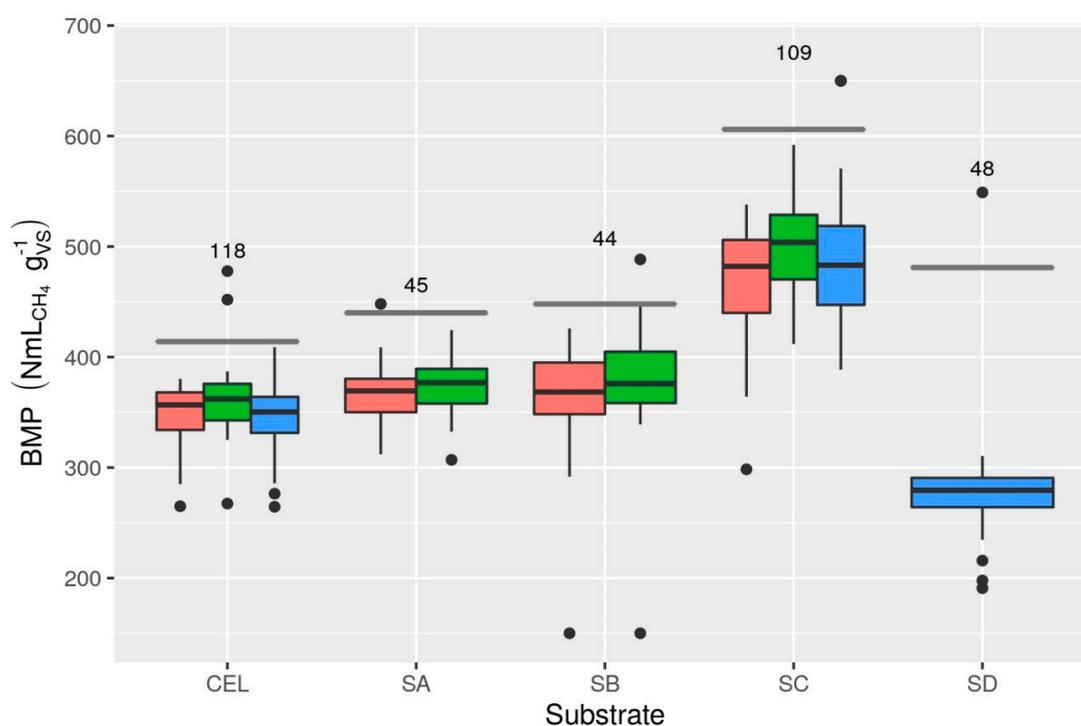


Figure 1. Boxplot summary of reported BMP values (subset B, mean values, $n = 3$), for three sets of tests: S1 T1 (red), S1 T2 (green), and S2 (blue). See Section 2.6.3 for boxplot description. Outliers were adjusted to a minimum of 150 and maximum of 650 NmL_{CH₄} g_{VS}⁻¹ for plotting (see SB and SC). Numeric plot labels show the number of BMP values for each substrate (all tests). Solid horizontal gray lines show theoretical maximum BMP (Table 1).

Calculated BMP values (subset E2, at the latest duration, using overall median substrate VS values, Table S3) generally showed similar variability as reported BMP measurements (Table 3). Where there were differences in RSD_R or RR, they were larger for calculated BMP, suggesting that differences might be due to data entry errors (see Section 3.2.2).

3.2. Evaluation of Sources of Error

Results described above (Section 3.1.2) show that reproducibility was generally far from the target, and that the presence of extreme values was a significant problem. Identifying sources of variability can help improve reproducibility. With raw laboratory measurements of CH₄ production as well as

VS measurements from each laboratory, it was possible to assess the importance of some individual sources of variability.

3.2.1. Volatile Solids Measurement

Systematic error in VS measurement translates into a proportional error in BMP. Variability in VS measurements was generally low, although there were a small number of extreme values (Figure S1). RSD in measured VS ranged from 1.4% for CEL VS in S1 to 4.1% for SC in S1 (Table S7). The largest RR for VS was 21%, much smaller than BMP RR. In general, VS variability was smaller than variability in reported BMP, suggesting that, for these homogeneous substrates at least, VS measurement error was not a major part of observed inter-laboratory variability. This result is also shown by comparison of BMP values calculated using individual participant-measured substrate VS values (Table S4) and median VS values (Table S3). Differences in RSD_R were generally small, and the use of median VS rarely substantially reduced RR. Clearly most inter-laboratory variability in BMP for these substrates is due to other sources. One possible exception is CEL, which was drier in S2 than in S1 (Figure S1). TS measurements were unusually variable in this case, including more low values. It is conceivable that water adsorption was more significant in S2, leading to uncertainty in the CEL VS mass added to BMP bottles (see Section 3.4.1). Alternatively, low reported VS values could reflect the inadvertent use of saved S1 material in S2, or simply the reuse of S1 VS measurements. It is important to note that VS measurement is likely a larger source of variability in BMP for some other substrates, particularly those with a significant volatile fraction, which require a correction [66–68].

3.2.2. Data Processing

Some differences in data processing calculations were apparent by directly comparing reported and calculated BMP (subsets A and C). The observed difference between reported and calculated BMP values could be due to deliberate differences in calculation methods but also errors in data entry and calculations. Differences were smallest (median of absolute values was 0.02% of mean BMP) for AMPTS II results (Figure 2). Because AMPTS II systems return CH_4 volume already standardized, BMP calculations are trivial, and in fact, only rounding error or other small differences (e.g., due to an inaccurate assumption of equal inoculum quantity in each bottle) are expected. Manometric and volumetric results showed larger differences (Figure 2). Median absolute values of differences were small, 2.6% for volumetric methods (excluding AMPTS II) and 3.4% for manual manometric methods. However, some differences were surprisingly large, exceeding 20% for some manometric measurements, and reaching 76% for a single volumetric result. Furthermore, apparent error varied among laboratories using the same methods. Standard deviation of relative calculation error was 4.0% for manometric and 13% for volumetric methods (relative to mean calculated BMP). The magnitude of error in many observations and, perhaps more importantly variation in this error, approached (manometric) or exceeded (volumetric) the magnitude of observed RSD_R in reported BMP values for several test by substrate combinations (Table 3).

These results suggest that calculation or data entry errors may play a significant role in inter-laboratory variability. However, the calculated data set (subset C, Table S2) actually showed more extreme values than reported (subset A, Table S1), and generally slightly higher RSD_R , highlighting the importance of small number of extreme and undoubtedly inaccurate BMP values. For these unusual observations, differences between calculated and reported results may be due to data entry mistakes. Regardless of the cause, these results show a troubling lack of verifiability in many BMP results.

A survey of participants provided some additional confirmation that calculation errors were indeed present. Participants mentioned omission of a correction for water vapor or using a different equation for the correction, evaluating CH_4 production at different times for blanks and bottles with substrate, use of a constant assumed biogas pressure instead of measured values for volume standardization, double counting of headspace CH_4 , and the use of different standard pressure (1 bar instead of 101.325 kPa). Reasons for the largest differences, however, were not provided.

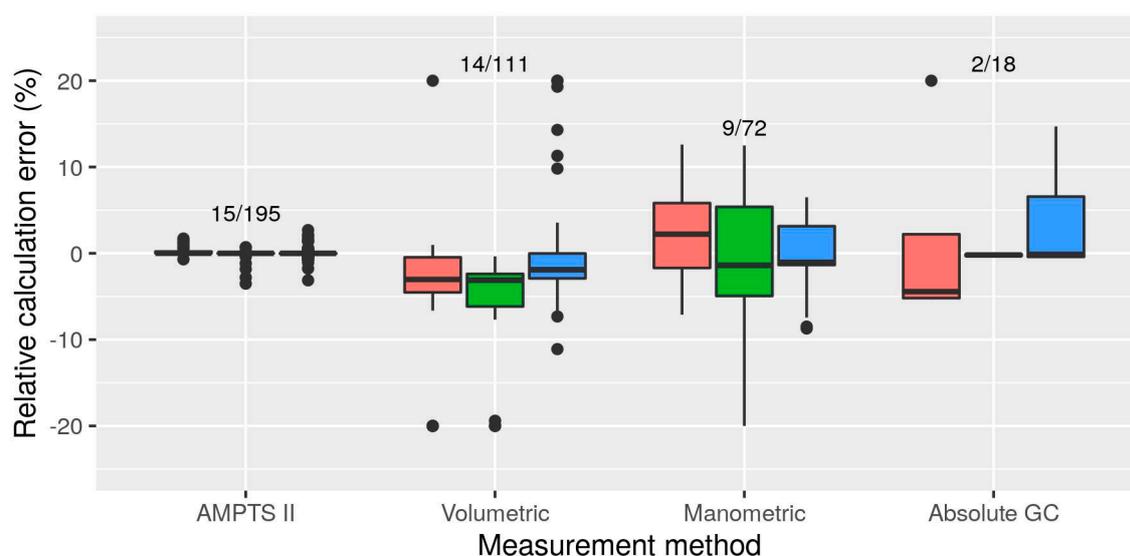


Figure 2. Difference between reported (subset A) and calculated (subset C) BMP values grouped by measurement method (relative to calculated BMP). A total of 10 volumetric observations and 1 absolute GC observation beyond $\pm 25\%$ were excluded from plot. Numeric labels show number of laboratories/BMP values for each method (both studies). Gravimetric results were excluded because reported BMP values (like calculated) were from the project organizers. Colors as in Figure 1.

Intra-test variability in BMP measurements is quantified by an RSD_B (based on a calculated standard error) associated with each BMP observation. Comparison of reported and calculated RSD_B values (S2 only) shows that even this calculation is not standardized, and many participants underestimated RSD_B substantially (Figure S2). A slight majority of laboratories (18 of 35 responding) only considered variability among bottles with substrate in their calculations, in contrast to the instructions provided to participants—that error from blanks should be included as well (Section 2.5). Most of the remainder included the two requested sources, but a small number of laboratories (3 of 35) included error from VS measurement as well (as in Section 2.5). Generally, variation in CH_4 production by bottles with substrate and inoculum was the largest source of error. The contribution of blanks was about half as large, and VS measurement variability contributed $< 10\%$ as much to RSD_B (median values). There were a small number of observations where these sources were much larger.

3.2.3. Inoculum Effects

There was virtually no evidence of consistent inoculum origin effects on BMP. Results from the first group of inoculum comparisons, where each lab used their regular inoculum and a single shared inoculum within each country (Section 2.4.2), did not show any reduction in BMP variability or a tendency for BMP to shift toward the mean when switching to a shared inoculum (Figure 3). In most cases, there was evidence of an increase in variability when using a shared inoculum (standard deviation among BMP values increased when switching to a shared inoculum for all combinations except CEL for country A ($p = 0.015$ from two-sample Wilcoxon test)). A mixed-effects model confirmed this conclusion: inoculum type (nested within country) was the smallest source of random error, and AIC was lower (better model) when it was excluded. This unexpected result might be an experimental artifact due to disruption of an otherwise healthy inoculum by transport and storage. Storage in particular can reduce BMP and should be minimized [30,69,70]. If this was the cause of the observed result, it was not a widespread effect, however, because in many cases BMP increased with the use of a shared inoculum. The second group of comparisons did not provide any clear evidence that inoculum origin was a major source of observed inter-laboratory variability either (Figure S3). ANOVA results provide insufficient evidence of overall inoculum source effects ($p = 0.15$ from F -test, Table S9).

In contrast, a laboratory effect was clear from the ANOVA, and it was much larger than the mean inoculum effect (mean square 3500 vs. 380 ($\text{NmL}_{\text{CH}_4} \text{gVS}^{-1}$)²).

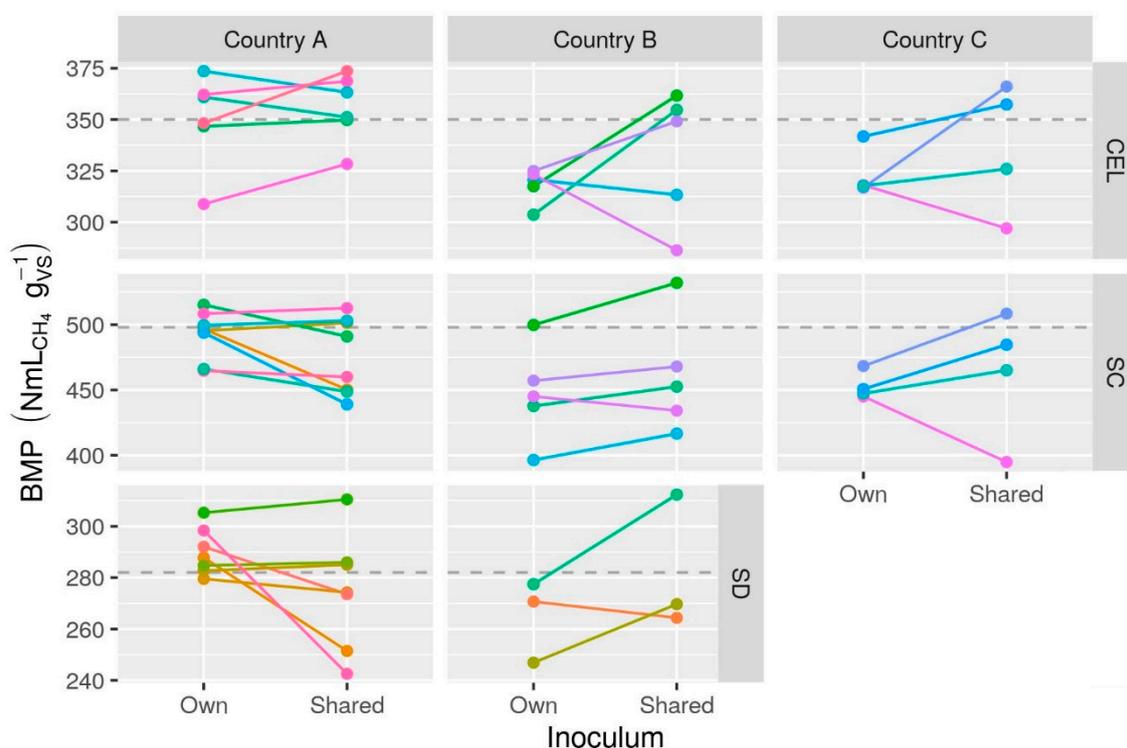


Figure 3. Apparent effects of inoculum origin on BMP from those laboratories that shared a single inoculum within each country (data subset G1, “Own” = regular inoculum source, which differed among laboratories, “Shared” = single inoculum shared within each country). Results from one particular laboratory were generally much higher than others and were excluded from the plots. Colors are unique for each combination of laboratory, measurement method, and ISR. Horizontal dashed lines show mean BMP for each substrate from data subset E1.

These comparisons might be expected to identify instances where a lab’s regular inoculum was insufficient, i.e., where BMP increased from an unusually low value into a reasonable range after switching. With the possible exception of CEL results from some Country B laboratories (Figure 3), this response was not apparent. These inoculum comparison results, in fact, seem to confirm that regular laboratory inocula were sufficient, and inoculum source was not a major contributor to observed inter-laboratory variability in BMP. In general, results from this large set of inoculum comparisons support studies that show small or no detectable differences in BMP measured with different inocula [23,25] and suggest that the large effects seen in some studies [24,26,29] may be due to inclusion of inocula with insufficient quality or the use of a fixed and insufficiently long test duration (Section 3.3).

A subset of laboratories used multiple ISRs for substrates SC and SD in one or more tests. These results support a general lack of substantial ISR effects, as long as ISR is sufficient (i.e., very low ISR values are expected to be problematic) (Figure S4). While there was some variation in BMP with ISR, it was generally small (<10%). This result supports the assumption of additivity used in calculating net CH_4 production [11,53]. Further evidence for additivity is found by the lack of any relationship between BMP and the fraction of total CH_4 production coming from inoculum (Figure S5). In fact, the limit of 20% stated by VDI [11] (p. 59) is not supported by these data, which do not show any obvious problems up to the maximum of ca. 40% (Figure S5). However, RSD_B and the magnitude of any possible non-additive effects both increase along with inoculum CH_4 production, so high values are not recommended.

3.2.4. Measurement Methods

A graphical assessment suggests that there were some consistent differences among measurement methods (Figure 4). Absolute GC measurements were typically much higher than others (or, in one case, lower). However, sample sizes were too small to explore these differences, and the absolute GC and gravimetric methods were excluded from a statistical comparison. For some substrates, there was a tendency for AMPTS II results to be higher than manometric or other volumetric results. Although differences appeared to vary among substrates and tests, there was some evidence of overall systematic differences among measurement methods in both S1 and S2 ($p < 0.003$ from ANOVA F -test). Manometric results were 10% lower than AMPTS II in S1 T2 ($p = 0.0016$ from Tukey’s HSD test) but not the other tests. In contrast, other volumetric results were lower than AMPTS II for both S1 T1 and S1 T2 ($p < 0.02$ from Tukey’s HSD test) by 14% and 8%, respectively. In contrast, other volumetric results were 10% higher than AMPTS II results in S2 ($p = 0.013$ from Tukey’s HSD test). Although there was evidence of consistent differences, most of the observed variability was unrelated to measurement method groups as shown by high variability within these groups (Figure 4).

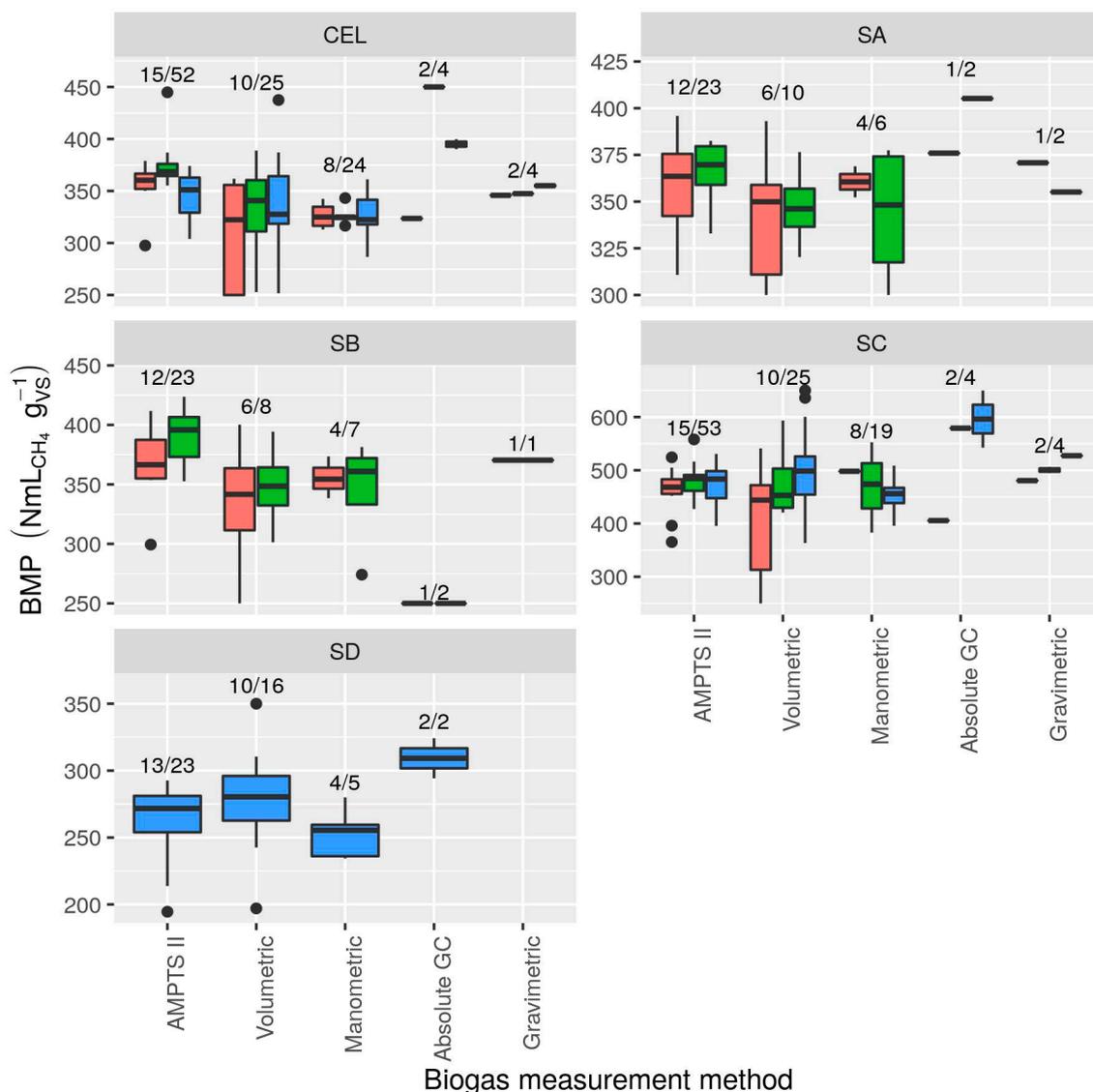


Figure 4. BMP vs measurement method (subset E1). Color as in Figure 1. Numeric labels show number of laboratories/BMP values for each method × substrate combination (total for both S1 and S2). Extreme values were adjusted for plotting (Section 2.6.3).

Widespread systematic error in BMP measurement may be indicated by correlation of substrate BMP to the BMP of a reference substrate, e.g., CEL. All three substrates showed moderate correlation (the non-parametric correlation coefficient Kendall's tau was 0.48–0.50, $p < 0.0001$ for all 4 substrates based on Kendall test) between BMP and the BMP of CEL measured in the same test (Figure 5). Slope estimates from robust regression ranged from 0.51 (SD) to 1.2 (SC), while a value of 1.0 is expected if variation were due to a simple systematic measurement bias. This result provides further evidence that differences among laboratories are at least partially systematic (consistent with the mixed-effects model results, Section 3.1.2), although clearly there are important random sources of error as well, considering the weakness of the correlation. The presence of clear correlation supports the use of stringent positive control validation criteria, because a tendency to measure high or low CEL BMP is reflected in the values measured for other substrates. However, these results do not support normalization or “correction” of substrate BMP measurements by a cellulose BMP result; variability in the correlations is simply too high, and there is no reason why large systematic errors in BMP measurement cannot be eliminated.

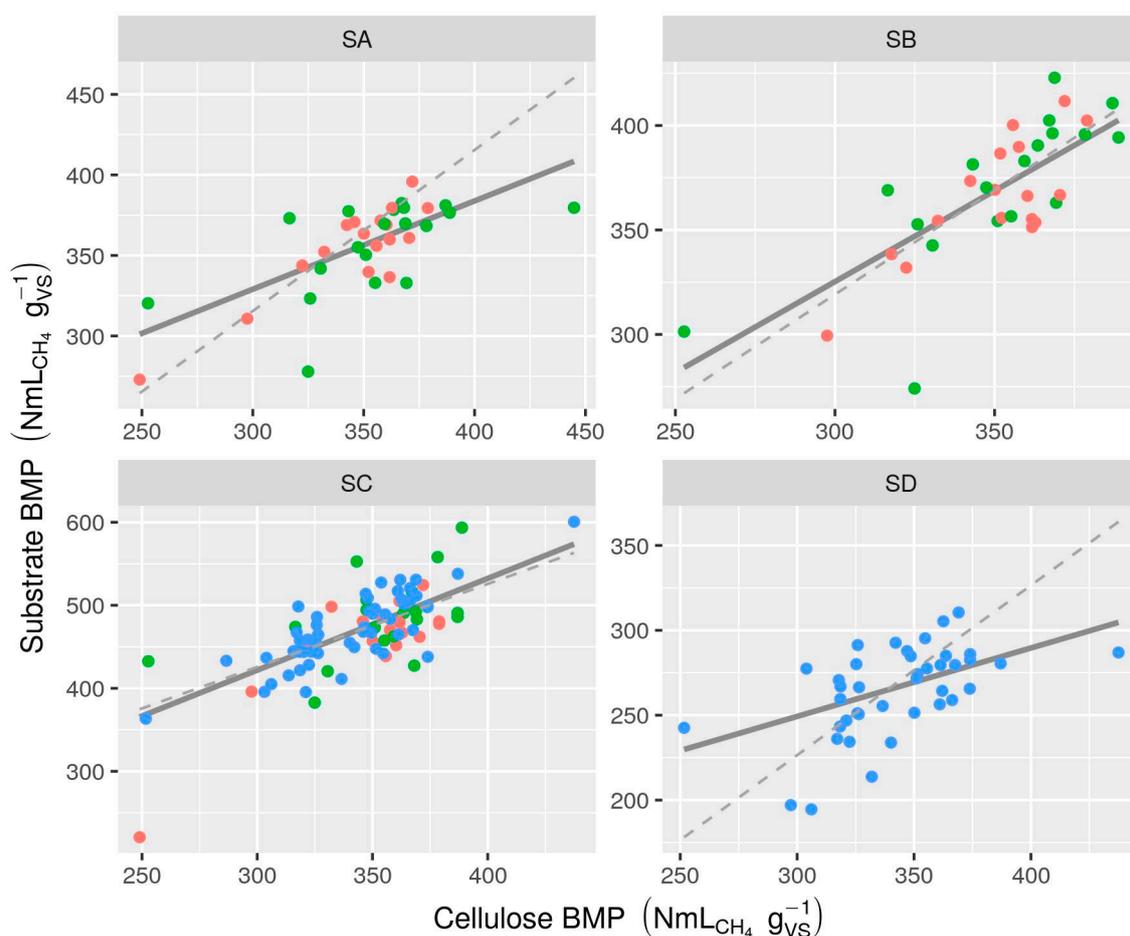


Figure 5. Substrate BMP versus cellulose BMP measured in the same test (subset E1). Colors as in Figure 1. Solid gray lines show robust regression result. Dashed lines have a slope of 1 and pass through median (included for slope comparison).

The addition of bottle weighing at the start and end of a subset of S2 BMP tests from 7 laboratories using manual methods provided an independent evaluation of measurement bias in these methods. Although low precision in measurement of small mass differences can lead to poor resolution in gravimetric results [12], the majority (88%) of observations suggest a negative bias in volume- or pressure-based methods (Figure 6). The mean relative apparent error (percentage of measured mass loss) based on the difference between measured and expected mass loss was 10% ($p = 0.00016$ based on

a one-sample *t*-test). Standard deviation of relative error among all observations was 32%, or 5.5% among mean values for each laboratory. This apparent bias alone is enough to explain a significant part of observed variability in BMP described above (Table 3), where about half the study \times test \times substrate groups had an RSD lower than 10%. Additionally, it is comparable to the differences in BMP among measurement methods (Figure 4), supporting the contention that systematic bias in some manual methods contributes to observed inter-laboratory variability in BMP. These results also highlight the value of adding gravimetric measurements to manual volumetric or manometric methods to check results.

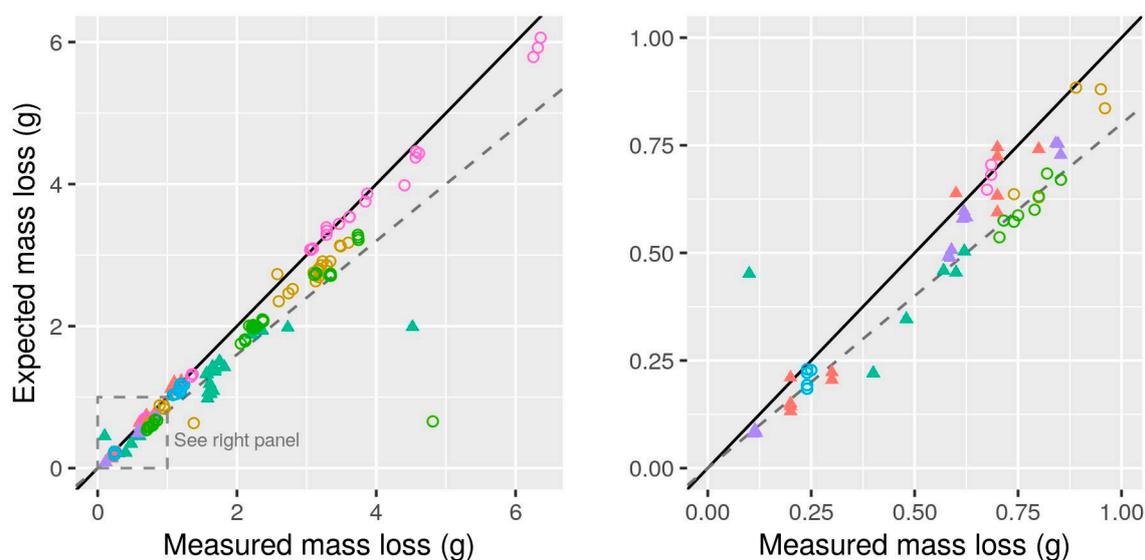


Figure 6. Apparent error in a subset of manual BMP measurements made by 7 laboratories shown by comparison of expected mass loss (calculated from total biogas volume over the entire BMP trial based on reported volume or pressure measurements) and actual mass loss (difference between initial and final bottle mass) (subset H). Right panel shows a close-up of same data shown on left (note axis limits). Both manual manometric (circles) and manual volumetric (triangles) measurements shown. Colors are unique for each laboratory. Dashed line shows -20% error.

Differences between measurement methods were apparent within laboratories as well (subset I, Section 2.6.2). The difference between two different automated volumetric systems used in one laboratory that tested CEL and SC, each with two different inocula, was about 5.5% (AMPTS II result higher, $p = 0.022$ from ANOVA *F*-test). Manual manometric BMP values were only slightly (average of 3.0%) lower than AMPTS II results in a set of tests from one laboratory that included CEL, SC, and SD ($p = 0.033$ from ANOVA *F*-test) (Figure S6), but manual volumetric results were nearly identical (1.4% lower). Lastly, a single laboratory measured BMP of CEL, SC, and SD using both manometric and gravimetric methods in a fully crossed experiment that included two different inocula. Here, BMP values were not clearly different (on average manometric results were $<1\%$ lower than gravimetric after dropping a single outlier, otherwise about 2% lower, but $p > 0.9$ from overall *F*-test). These differences, even when unambiguous, were small for the three laboratories, which is reflected in mixed-effects model estimates of method-based variability (as standard deviation) from this subset, which ranged from 0.2–3.6%, much lower than observed variability (Section 3.1.2). While differences may be present, it is possible to obtain similar or nearly identical BMP values using methods based on completely different principles, as has been shown previously in some cases [9,12,36]. Demonstration of large differences between methods [34–37] clearly indicates measurement errors and should not be accepted as unavoidable.

Although these results together seem complex, a single consistent explanation can explain them. Small biases (perhaps $< 15\%$) almost certainly exist in some measurement methods. However, most of

the error observed among laboratories is more likely to be due to particular details of measurement methods that are laboratory-, test-, or perhaps even technician-specific. Because these errors are not always associated with a simple method category (e.g., “manual manometric”), they may be difficult to detect in published measurements, highlighting the importance of validation criteria.

3.3. Evaluation of Duration Criteria

Most BMP tests were run to the requested duration: 93% of BMP tests (subset E1) were run to the 1% net 3 d criterion in S1 and 89% met the 0.5% net 3 d criterion in S2 (see Section 2.6.1 for description of duration criteria). Only 53% met the more stringent 0.5% gross 3 d criterion. As logically required, the 1% criterion duration was never larger than the more stringent 0.5% criterion. Furthermore, net criteria, based on net CH₄ production (after subtraction of the inoculum contribution), were almost always met before the corresponding gross (total) criteria, as expected. Therefore, the least stringent relative criterion considered was 1% net 3 d, and 0.5% gross 3 d was the most stringent, generally returning the longest durations. Comparing these two criteria showed some interesting trends (Figure 7), which were supported by a numeric summary, as described in the following. First, the duration of the relative criteria varied among labs and substrates. For most substrates, the 1% net duration was generally reached before 25 days, but for SD (expected to be the slowest-degrading substrate), it was usually later. Second, in most cases, differences in BMP at 1% net 3 d and 0.5% gross 3 d were small: median difference was 2.2% of the larger BMP values (25th and 75th percentiles: 0.6% and 3.6%). Third, differences in durations were generally large and sometimes very large; median difference was 12 days (25th and 75th percentiles: 5 and 16 days). These descriptions are not accurate for all observations. In rare cases, the difference between BMP for the two relative criteria was larger than 5%. A few cases provided some evidence of an improvement in BMP accuracy when comparing a fixed and relative duration; some of the 25 day BMP values that were relatively low were associated with a 1% net value closer to the median response (Figure S7). This improvement is what one might expect when switching to a relative duration, which overcomes effects of differences in kinetics, although there is little evidence of an overall reduction in inter-laboratory variability. 1% net 3 d durations were very short in some cases (<20 d, Figure 7) but associated BMP values were generally only slightly lower than those from the fixed duration (Figure S7).

These results show that the use of a relative duration eliminates excess incubation time, avoids insufficient incubation time for slowly degradable substrates (SD), and circumvents the challenge of identifying a single fixed duration that works in all cases. Furthermore, the 1% net 3 d criterion provides results similar to a very stringent 0.5% gross 3 d criterion but with much shorter incubation durations. Lastly, unlike gross criteria, this net criterion is independent of inoculum CH₄ production (assuming additivity), which might otherwise artificially extend durations.

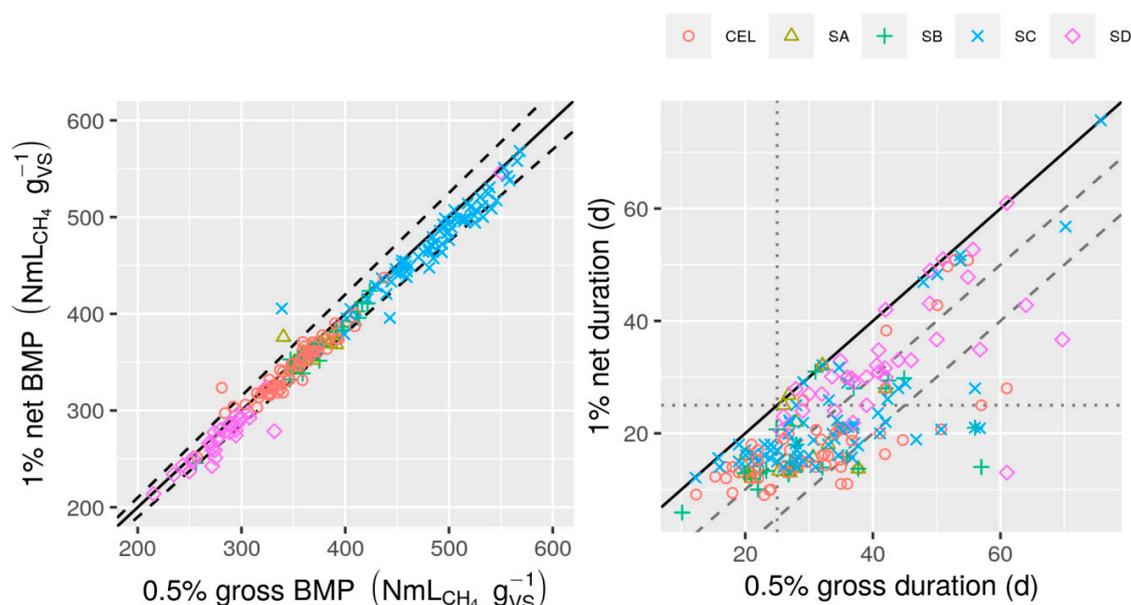


Figure 7. Comparison of BMP (**left**) and duration (**right**) based on the 0.5% gross 3 d and 1% net 3 d CH_4 production duration criteria (subset D). Observations include only those results where the test duration reached the more stringent criterion (0.5% gross 3 d), with nearly half (47%) of all observations omitted. Both studies included. Solid line shows 1:1 response, dashed lines $\pm 5\%$ (**left**) or +10 and 20 days (**right**), and dotted lines show 25 d (**right** only). Some outliers beyond axis limits were excluded.

3.4. Validation Criteria

A revised recommended set of validation criteria was developed through an evaluation of calculated BMP values and consideration of theory. In this section, the selection process is explained and evaluation results for three sets of validation criteria (Section 2.6.4) are presented.

3.4.1. Cellulose BMP Limits

Criteria based on cellulose BMP may eliminate BMP values with high error resulting from any number of reasons, including measurement errors, calculation errors, and inactive (or insufficiently diverse) inoculum. Results shown above (Section 3.2.4) suggest that this criterion is particularly important for reducing inter-lab variability. Without a precise “known” value for cellulose BMP, identification of the most accurate results and selection of validation criteria will always be somewhat arbitrary. The approach used here is based on both theory and BMP measurement.

Measurements of CEL BMP ranged from below 300 to above the theoretical maximum of $414 \text{ NmL}_{\text{CH}_4} \text{ g}_{\text{VS}}^{-1}$ (Figures 3 and 8). Theoretical calculations based on measured microbial yields suggest that about 85% of the theoretical maximum BMP ($352 \text{ NmL}_{\text{CH}_4} \text{ g}_{\text{VS}}^{-1}$) should typically be recovered if cellulose degradation is complete and no degradation of microbial biomass occurs [71,72]. The 1% net criterion does not guarantee complete degradation, but only ensures that BMP is probably near the maximum that would be reached in a longer incubation (Section 3.3). Conversely, microbial biomass degradation occurs during BMP trials, leading to a higher BMP, although the extent of degradation is difficult to predict and may vary among inocula. Adding to this uncertainty, measured values of parameters for microbial yields vary substantially [73,74], and resulting estimates of cellulose BMP (with no decay of microbial biomass) are as low as 60% [71]. However, the most extreme yields can be eliminated on the basis of energetics [48].

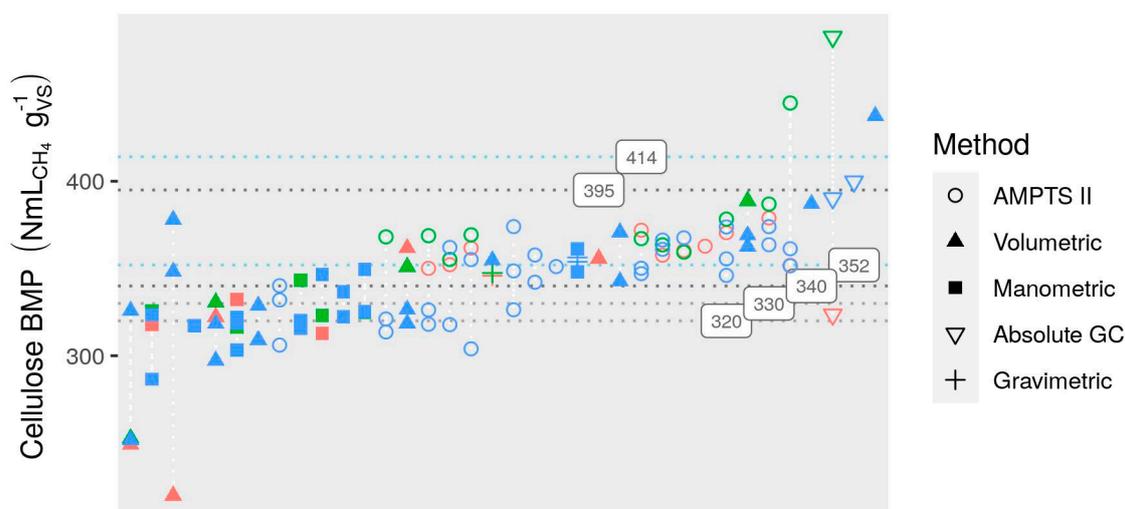


Figure 8. Calculated 1% net 3 d cellulose BMP results from all tests (subset E1), sorted in order of mean BMP. Dotted horizontal lines show the following limits: original [19] (blue), revision 2 (recommended) (dark gray), as well as 320 and 330 $\text{NmL}_{\text{CH}_4} \text{gVS}^{-1}$ (light gray). White dotted vertical lines connect results from individual labs. Colors indicate study and test, as in Figure 1.

A lower limit of 352 $\text{NmL}_{\text{CH}_4} \text{gVS}^{-1}$ eliminates many calculated BMP values (Figure 8). The lowest of these excluded values are from manometric and volumetric methods, which may have a tendency to be negatively biased (Section 3.2.4). Therefore, exclusion is probably appropriate. Other, higher, excluded observations are from AMPTS II, which is less likely to produce negatively biased results (Section 3.2.4). However, most of these excluded values were from labs that also provided much higher BMP values, and this poor intra-laboratory reproducibility implies a high likelihood of low accuracy for individual values, and it is therefore reasonable to eliminate these observations as well.

Gravimetric results are expected to have only small bias [12,36], and therefore, they provide a convenient reference point. Gravimetric methods (with gas analysis by GC) were used by only two laboratories in the present work, and results were close: 347 and 348 $\text{NmL}_{\text{CH}_4} \text{gVS}^{-1}$ for one and 357 and 360 $\text{NmL}_{\text{CH}_4} \text{gVS}^{-1}$ for the other. Recent gravimetric cellulose BMP results from two other laboratories are similar: 361 $\text{NmL}_{\text{CH}_4} \text{gVS}^{-1}$ [9] and 347 $\text{NmL}_{\text{CH}_4} \text{gVS}^{-1}$ [75]. (All these values are for the 1% net 3 d duration.) Variability among these four labs is low ($\text{RSD}_R < 2\%$) and the magnitude is plausible. Therefore, a lower limit slightly below 352 $\text{NmL}_{\text{CH}_4} \text{gVS}^{-1}$ and below these values was selected for the recommended validation criteria: 340 $\text{NmL}_{\text{CH}_4} \text{gVS}^{-1}$.

Any BMP value above the theoretical maximum BMP of 414 $\text{NmL}_{\text{CH}_4} \text{gVS}^{-1}$ is clearly inaccurate, although a limit might be slightly above this maximum to account for reasonable random error. However, few results approach this maximum, suggesting a lower value is possible, and therefore, a limit of 395 $\text{NmL}_{\text{CH}_4} \text{gVS}^{-1}$ was selected. This limit implies a minimum of about 5% of available cellulose electrons remain in non-degraded microbial biomass, which is plausible [71].

3.4.2. Random Error in Cellulose BMP (RSD_B)

Inclusion of cellulose RSD_B in a set of validation criteria increases the odds of flagging a wide range of problems, including, for example, leaking bottles and data recording errors. To maximize the probability of detecting problems that exist, as low an RSD_B limit as practical is desirable. Although BMP measurements depend both on biological activity and analytical measurements, results show that relatively low RSD_B is possible. CEL RSD_B calculated here (subset E1) was generally low (Figure S8), with a median of only 2.5%. Nearly 85% of all observations were below 6%. Therefore, for the recommended criteria, a cellulose RSD limit of 6% was proposed. As mentioned above (Section 2.5), these RSD_B values include three sources of error.

3.4.3. Random Error in Blanks and Substrate BMP

The original criteria [19] included an upper RSD limit for the SMP of inoculum (measured using blanks) of 5%. Variability in blanks is included in calculation of cellulose BMP RSD_B , and low inoculum SMP RSD may be difficult to attain for inocula with low CH_4 production rates, which is not expected to negatively impact BMP (in fact, ca. 50% of the results in set E1 did not meet the original 5% criterion). Furthermore, there was no clear relationship between this RSD and BMP (Figure S9), and therefore, it was not considered for inclusion in the revised criteria. However, elimination of this criterion makes it essential that RSD_B include variability in blanks (Section 3.5). A limit of 5% in RSD_B for homogenous substrates was also included in the original criteria [19]. Because of concerns that were expressed by participants of the Freising workshop that (1) a single universal value could not be identified for all substrates on the basis of measurements on homogeneous substrates (Section 2.3) and (2) that definition of “homogeneous” and “heterogeneous” as in the original criteria [19] was ambiguous, a limit was not included in the revised criteria. Furthermore, the limit on cellulose RSD_B is expected to generally identify problems with measurement precision.

3.4.4. Evaluation of Validation Criteria

Application of the three sets of validation criteria (Section 2.6.4) showed that all were effective in improving reproducibility, providing reductions in both RR and RSD_R (Figure 9, Table 4 and Tables S5 and S6). Not surprisingly, average BMP values tended to increase following application of validation criteria as low values were eliminated. The original and most stringent set (O) eliminated at least half of BMP observations for each substrate \times test combination, and 73% of all observations (Table 4). Still, this stringency did not provide proportional benefits with respect to reproducibility. Revisions 1 and 2 (R1 and R2) excluded far fewer observations (34% for revision 1, and 55% for revision 2), although RSD_R was similar for all three sets. RR for the original criteria and revision 2 was similar (Table 4), but higher for revision 1 (Figure 9, Table S6). Notable differences included S1 T2 SC, where the original criteria provided an RR of 4% while eliminating 70% of BMP observations. Revision 2 (R2) criteria eliminated only 35%, for a relatively large RR of 25% and an RSD_R of 7%, which were the maxima observed for R2. For S2 SC, however, revision 2 criteria provided much lower RSD_R and RR than the original set (due to exclusion of high CEL BMP values) but excluded fewer observations. RSD_R ranged from 3–7% and was $<6\%$ in most cases, and RR ranged from 9 to 25% for revision 2. The more lenient revision 1 (R1) provided only slightly lower mean BMP values than revision 2 (0–6% lower), but RSD_R was higher, ranging from 4–12%, with most values $>6\%$ (Table S6).

The dominant reason for rejection (non-validation) by revision 2 set was that cellulose BMP values were outside the limits, which resulted in the exclusion of 40% of observations (Figure 10). However, the cellulose BMP RSD_B limit of 6% excluded some values (23% of observations, half of which also failed the BMP value criterion), including several extreme values, demonstrating some utility (Figure 10).

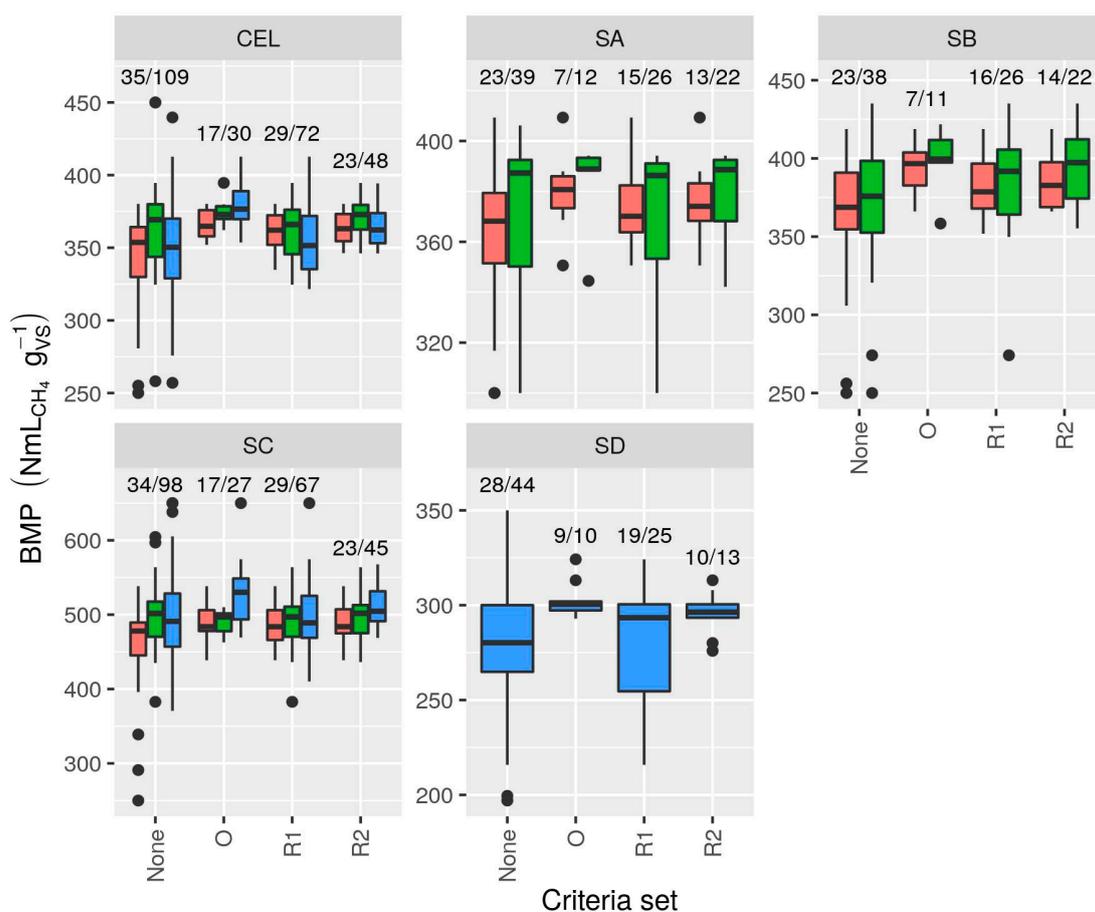


Figure 9. Effect of validation criteria (original (O) and revision 2 (R2)) application on resulting BMP values (subset E2, but excluding results from 7 BMP tests that did not include cellulose). Colors as in Figure 1. Numeric labels show number of laboratories/BMP values validated for each set. Extreme values were adjusted to plot near axis limits (Section 2.6.3).

Table 4. Numeric summary of validated calculated BMP values based on original (O) and revision 2 (R2) criteria sets (subset E2 but excluding results from 7 BMP tests that did not include cellulose). See Table 3 for additional notes.

Study	Test	Substrate	Validated * (%)		Mean BMP (NmLCH ₄ g _{VS} ⁻¹)		RSD _R (%)		RR (%)	
			O	R2	O	R2	O	R2	O	R2
S1	T1	CEL	41	55	367	363	2.9	3.1	8	9
S1	T1	SA	37	53	380	375	4.7	4.6	15	16
S1	T1	SB	32	53	394	385	4.8	4.8	13	14
S1	T1	SC	47	63	486	487	6.2	5.7	21	21
S1	T2	CEL	29	57	375	371	3.0	3.9	9	13
S1	T2	SA	25	60	382	378	5.5	5.3	13	13
S1	T2	SB	26	63	398	394	6.1	6.4	16	20
S1	T2	SC	30	65	490	501	3.8	7.1	10	25
S2	T1	CEL	23	36	380	365	4.5	3.6	16	13
S2	T1	SC	20	34	542	512	14.2	5.5	55	20
S2	T1	SD	23	30	303	296	3.1	3.4	10	13

* Percentage of observations (BMP values) validated by original (O) and revision 2 (R2) criteria sets.

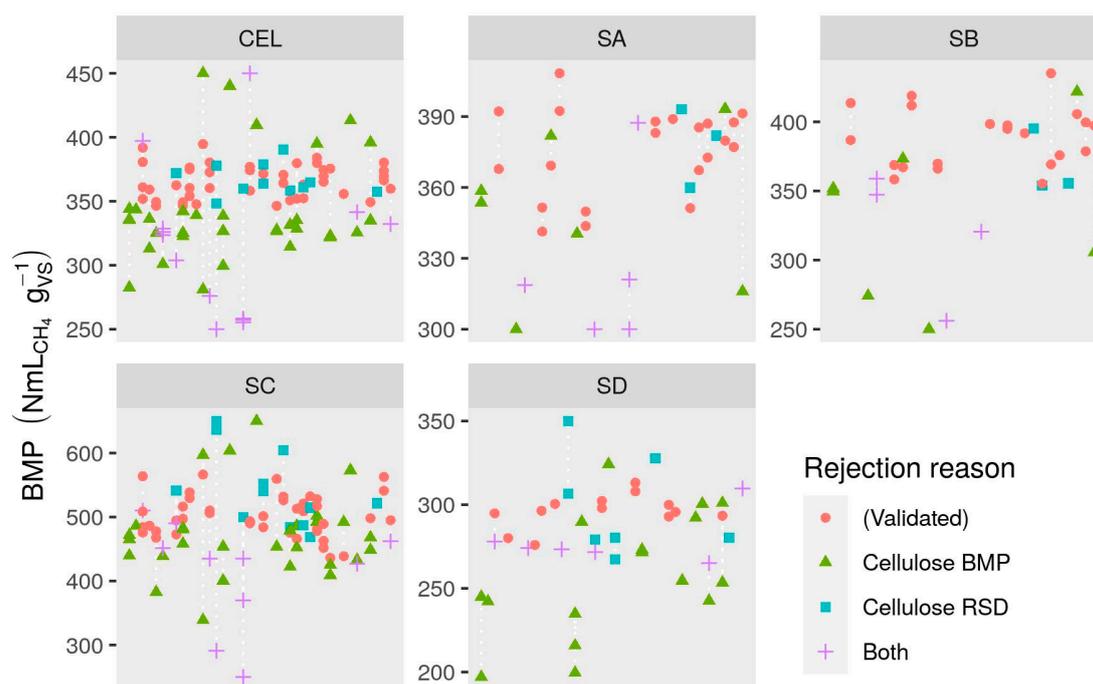


Figure 10. Reason for BMP rejection based on criteria set revision 2 (subset E2 but excluding results from 7 BMP tests that did not include cellulose). Position on x axis is random. White dotted vertical lines connect results from the same laboratory. Extreme values were adjusted to plot near axis limits (Section 2.6.3).

3.5. Recommendations

3.5.1. BMP Method Standardization

The results presented above show that BMP measurement suffers from a lack of standardization (Sections 3.1 and 3.2) but also that it is possible to improve reproducibility (Section 3.4.4). In this section recommendations for improving BMP reproducibility, based on results described above, are presented. These recommendations are intended for anyone making BMP measurements, including workers in both research and commercial applications. More details are available through a new website: <https://www.dbfz.de/en/BMP/>, aimed at improving the practice of BMP measurement. There, a more complete description of recommendations [76], detailed documentation of calculation methods [54–57], and other resources can be found. Standardization of a diverse set of laboratory and data processing methods should be expected to be an iterative process that requires input from the research community. Therefore, a transparent approach to document revision using GitHub has been included. Furthermore, the methods presented on this website need to be accepted by a large part of the research community if they are to truly become “standard” methods, and the site includes a mechanism for public approval of the documents. These documents may help address the problem of inconsistent BMP methods and lack of necessary detail in methods sections of papers [68].

3.5.2. BMP Measurement Methods

Results presented above confirm that measurement methods have biases. Therefore, assessment of measurement methods by BMP laboratories is important. Assessment can take different forms, including, very simply, the use of a positive control in every BMP trial, or informal inter-lab comparisons using cellulose in addition to more complex substrates. In the case of manual volumetric or manometric methods (which, while clearly able to provide accurate results, are apparently not always reliable), gravimetric measurements can easily be used to check accuracy [12,40]. In contrast to some earlier literature [34,35], differences among measurement methods should not be accepted as unavoidable or

indicative of effects on microbial activity. Differences in fact more likely show biases in one or more methods, and it is possible to eliminate these biases, as shown in the results of this study and in other studies [36,40].

3.5.3. Data Processing and Reporting

Surprisingly, calculations appear to be a significant source of error in BMP determination. Calculations should follow a common, accepted approach, and the detailed documentation now provided for free is strongly recommended [53–57]. Any departure from these methods should be clearly stated in publications. Custom templates (e.g., Excel) or scripts (e.g., Matlab or R) should be checked by comparing to standardized approaches. The web application the Online Biogas App (OBA) or the biogas package in R [45] can be used as a reference or even completely replace the use of custom data processing templates, as in recent publications [36,77]. Regardless, publications must clearly state how calculations were carried out and what software tools were used (see Section 2.6 for an example).

For the validation criteria described here, RSD_B calculations should include contributions from bottles with inoculum and substrate, blank samples, and VS measurements. Although the contribution of VS measurement variability was generally small here, this was not always the case. Inclusion of VS measurement variability will encourage careful and replicated VS measurements and reduce the risk of inaccurate BMP due to gross errors in TS/VS analysis.

Validation criteria set R2, based on cellulose BMP mean and RSD_B , should be applied to the results of all BMP tests. Any tests that do not meet all of the following criteria should be repeated:

- i. Duration at least 1% net 3 d (by substrate);
- ii. Cellulose BMP $RSD_B \leq 6\%$;
- iii. Cellulose BMP between 340 and 395 $NmL_{CH_4} g_{VS}^{-1}$.

When repeating a BMP test is not possible or practical, cellulose results and the lack of validation should be clearly stated in any report or results. However, it is acceptable to continue a BMP test beyond the 1% net 3 d duration in order to meet validation criteria. Triplicates ($n=3$) for each substrate and blanks are required for validation. Although apparent outliers may be eliminated if there is evidence of problems such as leakage or a gross error in setup, this should not be regularly done, and triplicates are still required after elimination of any bottles (therefore, $n > 3$ is a safer option). Reports and publications should include BMP and associated RSD_B for all tested substrates and cellulose, as well as test duration and the 1% net 3 d duration. BMP reported for the 1% net 3 d duration may be indicated as $BMP_{1\% \text{ net } 3d}$.

4. Summary and Conclusions

Analysis of BMP measurements from a large international effort showed that the use of a single protocol does not guarantee BMP values with low variability among laboratories. However, in combination with validation criteria, a standardized protocol can provide BMP values with high reproducibility (low inter-laboratory variability). This large project, unique in size and in the level of detail of collected data, provided results relevant to addressing the problem of high variability in BMP measurement. The most important of these, along with implications, are summarized below:

1. Even with the use of a single protocol, inter-laboratory variability was a significant problem, inflated by a small number of extreme values. Relative standard deviation among laboratories (RSD_R) was as high as 24%, and relative range 130%.
2. The validation criteria proposed by Holliger et al. [19], based on duration, mean cellulose BMP and variability in methane production from blanks, cellulose, and substrate were together effective in substantially reducing inter-laboratory variability. However, the majority of all BMP values were rejected by application of these criteria, including many that were apparently accurate.
3. Errors in data processing calculations or data entry (which are difficult to separate) were moderate, or in some cases, major sources of error in BMP. Additionally, calculation of relative standard

deviation for BMP values was done inconsistently among laboratories. Use of standardized approaches for data processing, as well as checking of calculations using standardized software, is strongly recommended.

4. There was evidence of differences among measurement methods even after re-calculation of all BMP values from original measurements: manual manometric and manual volumetric methods had a tendency to result in slightly (if not consistently) lower BMP values (as much as 14% below mean AMPTS II results). Evaluation of some manual methods based on mass loss measurement showed that negative bias was common (10% on average). Assessment of measurement biases, e.g., by comparing to gravimetric measurements, is recommended. Moderate correlation between substrate BMP and BMP of cellulose in the same test suggests that specific practices of laboratories or even technicians may be an underlying cause of observed inter-lab variability and that cellulose BMP is a promising indicator for validation. However, correlation is only moderate and cellulose results should not be used to “correct” potentially biased BMP measurements.
5. There was virtually no evidence of consistent effects of inoculum source on BMP, and any potential effects were much smaller than variation among labs. It is unlikely that inoculum source contributed substantially to observed inter-lab variation, suggesting that selection of a suitable inoculum was not a challenge for the participating laboratories. Large effects of inoculum source found in some other studies may be due to unusually ineffective inocula or insufficient duration and therefore may not be representative of typical BMP tests.
6. The best BMP duration criterion of those considered was when the daily net (after subtracting estimated inoculum contribution) CH₄ production (or production rate in mL d⁻¹) drops below 1% of cumulative net CH₄ production for at least 3 consecutive days (the “1% net 3 d” duration). Resulting BMP values were close to those from more stringent criteria, but duration was usually much shorter.
7. Based on evaluation of calculated BMP values, new validation criteria were proposed, and are recommended for use in all BMP tests:
 - a. Duration at least 1% net 3 d;
 - b. Cellulose RSD_B ≤ 6%;
 - c. Cellulose BMP between 340 and 395 NmL_{CH₄} g_{VS}⁻¹.

In total nearly half of all BMP values were validated, and many extreme values or those with likely negative bias were eliminated by application of these criteria. Inter-lab variability was also substantially improved, resulting in RSD_R < 8% and RR < 25% for all substrates and tests.

These results and, in particular, the recommendations, have the potential to substantially improve the quality of BMP measurements and therefore improve their value for both industry and research. However, any improvement depends on the widespread adoption of the proposed validation criteria. Current detailed recommendations and method documentation can be found at <https://www.dbfz.de/en/BMP>.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4441/12/6/1752/s1>, Figure S1: Substrate volatile solids (VS) measured by individual laboratories, Figure S2: Comparison of reported and calculated BMP RSD_B values, Figure S3: Apparent effects of inoculum origin on BMP, Figure S4: Calculated BMP versus ISR, Figure S5: Calculated BMP versus inoculum CH₄ production fraction, Figure S6: Calculated BMP versus measurement method for a single laboratory, Figure S7: Comparison of BMP based on a fixed 25 day duration and 1% d⁻¹ net CH₄ production criteria, Figure S8: Cellulose BMP relative standard deviation (RSD_B) for all calculated 1% net 3 d BMP values in subset E1, Figure S9: BMP vs. inoculum SMP RSD for calculated BMP, Table S1: Numeric summary of reported BMP values from subset A, Table S2: Numeric summary of calculated BMP values from subset C, Table S3: Numeric summary of calculated BMP values based on median substrate VS values at the latest available duration (subset E2), Table S4: Numeric summary of all calculated BMP values based on participant-measured substrate VS values (subset F), Table S5: Numeric summary of calculated BMP values from subset E2, but excluding results from 7 BMP tests that did not include cellulose, Table S6: Numeric summary of validated calculated BMP values based on revision 1 (R1) criteria sets, Table S7: Numeric summary of volatile solids measurements, Table S8: Restricted maximum likelihood estimates of random error sources in

reported BMP, Table S9: ANOVA results for evaluation of effect of inoculum origin on BMP (subset G2). Data entry templates: a compressed archive with the main data entry templates used in study S2. Data: a compressed archive with data sets and header descriptions as comma-delimited text files.

Author Contributions: Conceptualization, C.H. and H.F.d.L.; methodology, C.H., H.F.d.L., S.D.H., and K.K.; formal analysis, S.D.H.; investigation, C.H., H.F.d.L., K.K., and S.D.H.; resources, C.H., H.F.d.L., K.K., and S.D.H.; data curation, S.D.H. and C.H.; writing—original draft preparation, S.D.H. and C.H.; writing—review and editing, S.D.H., K.K., C.H. and H.F.d.L.; visualization, S.D.H.; supervision, C.H. and S.D.H.; project administration, C.H., H.F.d.L., K.K. and S.D.H.; funding acquisition, C.H. and K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Swiss Federal Office for Energy, which financed the management part of the both inter-laboratory studies as well as the participation in AD15 and AD16 conferences. The financial support of the TUM Global Incentive Fund to host the second international workshop on standardization of BMP tests in Freising (Germany) is highly appreciated.

Acknowledgments: All people and laboratories that participated in the inter-laboratory studies are acknowledged for having kindly carried out the tests for free. They are Madalena Alves and João Vítor Oliveira, Centre of Biological Engineering, University of Minho, Braga, Portugal; Irini Angelidaki and Ioannis Fotidis, Technical University of Denmark, Lyngby, Denmark; Sergi Astals, Advanced Water Management Center, The University of Queensland, Brisbane, Australia; Samet Azman and Lise Appels, KU Leuven, PETLab, Sint-Katelijne-Wave, Belgium; Alexandre Bagnoud, HEIG-VD, Yverdon-les-Bains, Switzerland; Urs Baier and Judith Krautwald, Institute for Chemistry and Biotechnology, ZHAW School of Life Sciences and Facility Management, Wädenswil, Switzerland; Yadira Bajon Fernandez, Cranfield University, Bedfordshire, United Kingdom; Alexander Bauer and Javier Lizasoain, University of Natural Resources and Life Sciences, Vienna, Austria; David Bolzonella and Federico Battista, University of Verona, Italy; Claire Bougrier, VERI (Veolia), Limay, France; Camilla Braguglia and Agata Gallipoli, IRSA-CNR, Monterotondo, Italy; Pierre Buffière, Université de Lyon, INSA-Lyon, Lyon, France; Marta Carballa and Anton Taboada, Department of Chemical Engineering, Institute of Technology, Universidade de Santiago de Compostela, Santiago de Compostela, Spain; Vasilis Dandikas and Diana Andrade, Bavarian State Research Center for Agriculture, Freising, Germany; Belén Fernández, IRTA, Barcelona, Spain; Elena Ficara, Arianna Catenacci, and Isabella Porqueddu, Politecnico di Milano-DICA, Milano, Italy; Jean-Claude Frigon, National Research Council Canada, Montréal, Canada; Jörn Heerenklage, Hamburg University of Technology, Hamburg, Germany; Ilona Sarvari Horvath, The Swedish Centre for Resource Recovery, University of Borås, Borås, Sweden; Pavel Jeníček and Jan Bartáček, University of Chemistry and Technology Prague, Prague, Czech Republic; Earl Jenson and Sylvaus Ekwe, Alberta Innovates—Technology Futures, Vegreville, Canada; K.K., Chair of Urban Water Systems Engineering, Technical University of Munich, Garching, Germany; Jing Liu and Mihaela Nistor, Bioprocess Control AB, Lund, Sweden; Rosa Marchetti and Ciro Vasmara, Research Centre for Animal Production and Aquaculture, Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, San Cesario (Modena), Italy; Florian Monlau and Christine Peyrelasse, APESA, Lescar, France; Hans Oechsner and Benedikt Hülsemann, State Institute of Agricultural Engineering and Bioenergy, University of Hohenheim, Stuttgart, Germany; André Pauss, Alliance Sorbonne Universités, UTC, Compiègne, France; Sébastien Pommier, TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France; Francisco Raposo, Instituto de la Grasa, Consejo Superior de Investigaciones Científicas, Seville, Spain; Thierry Ribeiro, Laura André, Institut Polytechnique UniLaSalle Beauvais, EA 7519 Transformations & Agroressources, Beauvais, France; Christian Schaum, Universität der Bundeswehr München, Institut für Wasserwesen, Neubiberg, Germany; Sebastian Schwede, Mälardalens högskola, Västerås, Sweden; Mariangela Soldano, Centro Ricerche Produzioni Animali, Reggio Emilia, Italy; Michel Torrijos and Romain Cresson, INRA, Laboratoire de Biotechnologie de l'Environnement, Narbonne, France; Miriam van Eekert and Els Schuman, LeAF, Wageningen, The Netherlands; Jules van Lier and Ralph Lindeboom, Delft University of Technology, Delft, The Netherlands; Harald Wedwitschka, Peter Fischer and Sören Weinrich, DBFZ Deutsches Biomasseforschungszentrum, Leipzig, Germany; Isabella Wierinck and Fabian de Wilde, OWS nv, Gent, Belgium. Sören Weinrich is acknowledged for a helpful discussion on microbial biomass yields, as well as help with website development. Website hosting by DBFZ Deutsches Biomasseforschungszentrum is gratefully acknowledged. Finally, we are thankful to Michel Roulin from Cargill Feed & Nutrition Switzerland who provided three substrates for free and to the people from the Bavarian State Research Center for Agriculture, Freising, Germany, who provided the new substrate rich in lignocellulose.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Carrere, H.; Antonopoulou, G.; Affes, R.; Passos, F.; Battimelli, A.; Lyberatos, G.; Ferrer, I. Review of feedstock pretreatment strategies for improved anaerobic digestion: From lab-scale research to full-scale application. *Bioresour. Technol.* **2016**, *199*, 386–397. [[CrossRef](#)] [[PubMed](#)]
2. Ariunbaatar, J.; Panico, A.; Esposito, G.; Pirozzi, F.; Lens, P.N.L. Pretreatment methods to enhance anaerobic digestion of organic solid waste. *Appl. Energy* **2014**, *123*, 143–156. [[CrossRef](#)]

3. Holliger, C.; Fruteau de Lacos, H.; Hack, G. Methane production of full-scale anaerobic digestion plants calculated from substrate's biomethane potentials compares well with the one measured on-site. *Front. Energy Res.* **2017**, *5*. [[CrossRef](#)]
4. Li, C.; Nges, I.A.; Lu, W.; Wang, H. Assessment of the degradation efficiency of full-scale biogas plants: A comparative study of degradation indicators. *Bioresour. Technol.* **2017**, *244*, 304–312. [[CrossRef](#)] [[PubMed](#)]
5. Lindorfer, H.; Pérez López, C.; Resch, C.; Braun, R.; Kirchmayr, R. The impact of increasing energy crop addition on process performance and residual methane potential in anaerobic digestion. *Water Sci. Technol.* **2007**, *56*, 55–63. [[CrossRef](#)] [[PubMed](#)]
6. Ruile, S.; Schmitz, S.; Mönch-Tegeder, M.; Oechsner, H. Degradation efficiency of agricultural biogas plants—A full-scale study. *Bioresour. Technol.* **2015**, *178*, 341–349. [[CrossRef](#)]
7. Owen, W.F.; Stuckey, D.C.; Healy, J.B., Jr.; Young, L.Y.; McCarty, P.L. Bioassay for monitoring biochemical methane potential and anaerobic toxicity. *Water Res.* **1979**, *13*, 485–492. [[CrossRef](#)]
8. Hansen, T.L.; Schmidt, J.E.; Angelidaki, I.; Marca, E.; Jansen, J.L.C.; Mosbæk, H.; Christensen, T.H. Method for determination of methane potentials of solid organic waste. *Waste Manag.* **2004**, *24*, 393–400. [[CrossRef](#)]
9. Justesen, C.G.; Astals, S.; Mortensen, J.R.; Thorsen, R.; Koch, K.; Weinrich, S.; Triolo, J.M.; Hafner, S.D. Development and validation of a low-cost gas density method for measuring biochemical methane potential (bmp). *Water* **2019**, *11*, 2431. [[CrossRef](#)]
10. Pabón Pereira, C.P.; Castañares, G.; Van Lier, J.B. An OxiTop[®] protocol for screening plant material for its biochemical methane potential (BMP). *Water Sci. Technol.* **2012**, *66*, 1416–1423. [[CrossRef](#)]
11. Verein Deutsch er Ingenieure e.V. *Fermentation of Organic Materials: Characterisation of the Substrate, Sampling, Collection of Material Data, Fermentation Tests*; Verein Deutsch er Ingenieure e.V.: Düsseldorf, Germany, 2016.
12. Hafner, S.D.; Rennuit, C.; Triolo, J.M.; Richards, B.K. Validation of a simple gravimetric method for measuring biogas production in laboratory experiments. *Biomass Bioenergy* **2015**, *83*, 297–301. [[CrossRef](#)]
13. Strömberg, S.; Nistor, M.; Liu, J. Towards eliminating systematic errors caused by the experimental conditions in Biochemical Methane Potential (BMP) tests. *Waste Manag.* **2014**, *34*, 1939–1948. [[CrossRef](#)] [[PubMed](#)]
14. Cresson, R.; Pommier, S.; Beline, F.; Bouchez, T.; Bougrier, C.; Buffière, P.; Pauss, A.; Pouech, P.; Preys, S.; Ribeiro, T. Results from a French Inter-Laboratory Campaign on the Biological Methane Potential of Solid Substrates. Available online: <https://hal.archives-ouvertes.fr/hal-01947463> (accessed on 17 June 2020).
15. Raposo, F.; Fernandez-Cegri, V.; De la Rubia, M.A.; Borja, R.; Beline, F.; Cavinato, C.; Demirer, G.; Fernandez, B.; Fernandez-Polanco, M.; Frigon, J.C.; et al. Biochemical methane potential (BMP) of solid organic substrates: Evaluation of anaerobic biodegradability using data from an international interlaboratory study. *J. Chem. Technol. Biotechnol.* **2011**, *86*, 1088–1098. [[CrossRef](#)]
16. Weinrich, S.; Schäfer, F.; Liebetrau, J.; Bochmann, G.; Paterson, M.; Oechsner, H.; Tillmann, P. *Value of Batch Tests for Biogas Potential Analysis: Method Comparison and Challenges of Substrate and Efficiency Evaluation of Biogas Plants*; IEA Bioenergy: Paris, France, 2018; ISBN 978-1-910154-49-6.
17. Angelidaki, I.; Alves, M.; Bolzonella, D.; Borzacconi, L.; Campos, J.L.; Guwy, A.J.; Kalyuzhnyi, S.; Jenicek, P.; Van Lier, J.B. Defining the biomethane potential (BMP) of solid organic wastes and energy crops: A proposed protocol for batch assays. *Water Sci. Technol.* **2009**, *59*, 927–934. [[CrossRef](#)]
18. Verein Deutsch er Ingenieure e.V. *Fermentation of Organic Materials*; Verein Deutsch er Ingenieure e.V.: Düsseldorf, Germany, 2006.
19. Holliger, C.; Alves, M.; Andrade, D.; Angelidaki, I.; Astals, S.; Baier, U.; Bougrier, C.; Buffière, P.; Carballa, M.; De Wilde, V.; et al. Towards a standardization of biomethane potential tests. *Water Sci. Technol.* **2016**, *74*, 2515–2522. [[CrossRef](#)] [[PubMed](#)]
20. Filer, J.; Ding, H.H.; Chang, S. Biochemical Methane Potential (BMP) Assay Method for Anaerobic Digestion Research. *Water* **2019**, *11*, 921. [[CrossRef](#)]
21. Pearse, L.F.; Hettiaratchi, J.P.; Kumar, S. Towards developing a representative biochemical methane potential (BMP) assay for landfilled municipal solid waste—A review. *Bioresour. Technol.* **2018**, *254*, 312–324. [[CrossRef](#)]
22. Raposo, F.; De la Rubia, M.A.; Fernandez-Cegri, V.; Borja, R. Anaerobic digestion of solid organic substrates in batch mode: An overview relating to methane yields and experimental procedures. *Renew. Sustain. Energy Rev.* **2012**, *16*, 861–877. [[CrossRef](#)]
23. Koch, K.; Lippert, T.; Drewes, J.E. The role of inoculum's origin on the methane yield of different substrates in biochemical methane potential (BMP) tests. *Bioresour. Technol.* **2017**, *243*, 457–463. [[CrossRef](#)]

24. De Vrieze, J.; Raport, L.; Willems, B.; Verbrugge, S.; Volcke, E.; Meers, E.; Angenent, L.T.; Boon, N. Inoculum selection influences the biochemical methane potential of agro-industrial substrates. *Microb. Biotechnol.* **2015**, *8*, 776–786. [CrossRef]
25. Hülsemann, B.; Zhou, L.; Merkle, W.; Hassa, J.; Müller, J.; Oechsner, H. Biomethane Potential Test: Influence of Inoculum and the Digestion System. *Appl. Sci.* **2020**, *10*, 2589. [CrossRef]
26. Dechrugsa, S.; Kantachote, D.; Chaiprapat, S. Effects of inoculum to substrate ratio, substrate mix ratio and inoculum source on batch co-digestion of grass and pig manure. *Bioresour. Technol.* **2013**, *146*, 101–108. [CrossRef] [PubMed]
27. Reilly, M.; Dinsdale, R.; Guwy, A. The impact of inocula carryover and inoculum dilution on the methane yields in batch methane potential tests. *Bioresour. Technol.* **2016**, *208*, 134–139. [CrossRef] [PubMed]
28. Hagen, L.H.; Vivekanand, V.; Pope, P.B.; Eijsink, V.G.H.; Horn, S.J. The effect of storage conditions on microbial community composition and biomethane potential in a biogas starter culture. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 5749–5761. [CrossRef] [PubMed]
29. Elbeshbishy, E.; Nakhla, G.; Hafez, H. Biochemical methane potential (BMP) of food waste and primary sludge: Influence of inoculum pre-incubation and inoculum source. *Bioresour. Technol.* **2012**, *110*, 18–25. [CrossRef]
30. Astals, S.; Koch, K.; Weinrich, S.; Hafner, S.D.; Tait, S.; Peces, M. Impact of Storage Conditions on the Methanogenic Activity of Anaerobic Digestion Inocula. *Water* **2020**, *12*, 1321. [CrossRef]
31. Fabbri, A.; Serranti, S.; Bonifazi, G. Biochemical methane potential (BMP) of artichoke waste: The inoculum effect. *Waste Manag. Res.* **2014**, *32*, 207–214. [CrossRef]
32. Raposo, F.; Banks, C.J.; Siegert, I.; Heaven, S.; Borja, R. Influence of inoculum to substrate ratio on the biochemical methane potential of maize in batch tests. *Process Biochem.* **2006**, *41*, 1444–1450. [CrossRef]
33. Rodriguez-Chiang, L.M.; Dahl, O.P. Effect of inoculum to substrate ratio on the methane potential of microcrystalline cellulose production wastewater. *BioResources* **2015**, *10*, 898–911. [CrossRef]
34. Himanshu, H.; Voelklein, M.A.; Murphy, J.D.; Grant, J.; O’Kiely, P. Factors controlling headspace pressure in a manual manometric BMP method can be used to produce a methane output comparable to AMPTS. *Bioresour. Technol.* **2017**, *238*, 633–642. [CrossRef]
35. Valero, D.; Montes, J.A.; Rico, J.L.; Rico, C. Influence of headspace pressure on methane production in Biochemical Methane Potential (BMP) tests. *Waste Manag.* **2016**, *48*, 193–198. [CrossRef] [PubMed]
36. Hafner, S.D.; Astals, S. Systematic error in manometric measurement of biochemical methane potential: Sources and solutions. *Waste Manag.* **2019**, *91*, 147–155. [CrossRef] [PubMed]
37. Kleinheinz, G.; Hernandez, J. Comparison of two laboratory methods for the determination of biomethane potential of organic feedstocks. *J. Microbiol. Methods* **2016**, *130*, 54–60. [CrossRef] [PubMed]
38. Pham, C.H.; Triolo, J.M.; Cu, T.T.T.; Pedersen, L.; Sommer, S.G. Validation and recommendation of methods to measure biogas production potential of animal manure. *Asian Australas. J. Anim. Sci.* **2013**, *26*, 864–873. [CrossRef] [PubMed]
39. Richards, B.K.; Cummings, R.J.; White, T.E.; Jewell, W.J. Methods for kinetic-analysis of methane fermentation in high solids biomass digesters. *Biomass Bioenergy* **1991**, *1*, 65–73. [CrossRef]
40. Hafner, S.D.; Rennuit, C.; Olsen, P.J.; Pedersen, J.M. Quantification of leakage in batch biogas assays. *Water Pract. Technol.* **2018**, *13*, 52–61. [CrossRef]
41. Svensson, K.; Kjørlaug, O.; Horn, S.J.; Agger, J.W. Comparison of approaches for organic matter determination in relation to expression of bio-methane potentials. *Biomass Bioenergy* **2017**, *100*, 31–38. [CrossRef]
42. Fanelli, D. Do pressures to publish increase scientists’ bias? An empirical support from US states data. *PLoS ONE* **2010**, *5*. [CrossRef]
43. International Organization for Standardization. *Solid Biofuels—Determination of Total Content of Carbon, Hydrogen and Nitrogen (iso 16948:2015)*; International Organization for Standardization: Geneva, Switzerland, 2015.
44. International Organization for Standardization. *Solid Biofuels—Conversion of Analytical Results from One Basis to Another (iso 16993:2016)*; International Organization for Standardization: Geneva, Switzerland, 2016.
45. Hafner, S.D.; Koch, K.; Carrere, H.; Astals, S.; Weinrich, S.; Rennuit, C. Software for biogas research: Tools for measurement and prediction of methane production. *SoftwareX* **2018**, *7*, 205–210. [CrossRef]
46. Hafner, S.; Rennuit, C.; Justesen, C.G.; Løjborg, N.; Mortensen, J.R.; Biogas Package v. 1.24.3. Available online: <https://github.com/sashahafner/biogas> (accessed on 7 April 2020).

47. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
48. Rittmann, B.E.; McCarty, P.L. *Environmental Biotechnology: Principles and Applications*; McGraw-Hill: Boston, MA, USA, 2001; ISBN 0-07-234553-5.
49. Environmental Protection Agency. *Method 1684 Total, Fixed, and Volatile Solids in Water, Solids, and Biosolids*; U.S. Environmental Protection Agency, Office of Water, Office of Science and Technology Engineering and Analysis Division (4303): Washington, DC, USA, 2001.
50. American Public Health Association, American Water Works Associate, and Water Environment Federation. *Standard Methods for the Examination of Water and Wastewater*, 21st ed.; APHA-AWWA-WEF: Washington, DC, USA, 2005; ISBN 978-0-87553-047-5.
51. BioProcess Control. AMPTS II—Methane Potential Analysis Tool. Available online: <https://www.bioprocesscontrol.com/products/ampts-ii> (accessed on 19 June 2020).
52. Rozzi, A.; Remigi, E. Methods of assessing microbial activity and inhibition under anaerobic conditions: A literature review. *Rev. Environ. Sci. Biotechnol.* **2004**, *3*, 93–115. [[CrossRef](#)]
53. Hafner, S.D.; Astals, S.; Holliger, C.; Koch, K.; Weinrich, S. Calculation of Biochemical Methane Potential (BMP). Standard BMP Methods Document 200, Version 1.6. Available online: <https://www.dbfz.de/en/BMP> (accessed on 19 April 2020).
54. Hafner, S.D.; Løjborg, N.; Astals, S.; Holliger, C.; Koch, K.; Weinrich, S. Calculation of Methane Production from Volumetric Measurements. Standard BMP Methods Document 201, Version 1.5. Available online: <https://www.dbfz.de/en/BMP> (accessed on 19 April 2020).
55. Hafner, S.D.; Astals, S.; Buffiere, P.; Løjborg, N.; Holliger, C.; Koch, K.; Weinrich, S. Calculation of Methane Production from Manometric Measurements. Standard BMP Methods Document 202, Version 2.5. Available online: <https://www.dbfz.de/en/BMP> (accessed on 19 April 2020).
56. Hafner, S.D.; Richards, B.K.; Astals, S.; Holliger, C.; Koch, K.; Weinrich, S. Calculation of Methane Production from Gravimetric Measurements. Standard BMP Methods Document 203, Version 1.0. Available online: <https://www.dbfz.de/en/BMP> (accessed on 19 April 2020).
57. Hafner, S.D.; Justesen, C.; Thorsen, R.; Astals, S.; Holliger, C.; Koch, K.; Weinrich, S. Calculation of Methane Production from Gas Density-Based Measurements. Standard BMP Methods Document 204, Version 1.5. Available online: <https://www.dbfz.de/en/BMP> (accessed on 19 April 2020).
58. Crowder, M. Interlaboratory comparisons: Round robins with random effects. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1992**, *41*, 409–425. [[CrossRef](#)]
59. Zar, J.H. *Biostatistical Analysis*, 4th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1999; ISBN 0-13-081542-X.
60. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [[CrossRef](#)]
61. Faraway, J.J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, 1st ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2006; ISBN 1-58488-424-X.
62. Faraway, J.J. *Linear Models with R*; Texts in statistical science; Chapman & Hall/CRC: Boca Raton, FL, USA, 2005; ISBN 1-58488-425-8.
63. Rosenbaum, P.R. *Design of Observational Studies*; Springer: Berlin/Heidelberg, Germany, 2009; ISBN 978-1-4419-1213-8.
64. Bauer, D.F. Constructing Confidence Sets Using Rank Statistics. *J. Am. Stat. Assoc.* **1972**, *67*, 687–690. [[CrossRef](#)]
65. Venables, W.N. *Modern Applied Statistics with S Statistics and Computing*, 4th ed.; Springer: New York, NY, USA, 2002; ISBN 0-387-95457-0.
66. Kreuger, E.; Nges, I.A.; Bjornsson, L. Ensiling of crops for biogas production: Effects on methane yield and total solids determination. *Biotechnol. Biofuels* **2011**, *4*, 44. [[CrossRef](#)]
67. Weissbach, F.; Strubelt, C. Correcting the dry matter content of maize silages as a substrate for biogas production. *Landtechnik* **2008**, *63*, 82–83.
68. Raposo, F.; Borja, R.; Ibelli-Bianco, C. Predictive regression models for biochemical methane potential tests of biomass samples: Pitfalls and challenges of laboratory measurements. *Renew. Sustain. Energy Rev.* **2020**, *127*, 109890. [[CrossRef](#)]
69. Wang, B.; Strömberg, S.; Nges, I.A.; Nistor, M.; Liu, J. Impacts of inoculum pre-treatments on enzyme activity and biochemical methane potential. *J. Biosci. Bioeng.* **2016**, *121*, 557–560. [[CrossRef](#)] [[PubMed](#)]

70. Koch, K.; Hafner, S.D.; Weinrich, S.; Astals, S. Identification of critical problems in biochemical methane potential (BMP) tests from methane production curves. *Front. Environ. Sci.* **2019**, *7*, 178. [[CrossRef](#)]
71. Weinrich, S. *Praxisnahe Modellierung von Biogasanlagen: Systematische Vereinfachung des Anaerobic Digestion Model No. 1 (ADM1)*; Universität Rostock: Rostock, Germany, 2018.
72. Weinrich, S.; Nelles, M. Critical comparison of different model structures for the applied simulation of the anaerobic digestion of agricultural energy crops. *Bioresour. Technol.* **2015**, *178*, 306–312. [[CrossRef](#)]
73. Batstone, D.J.; Keller, J.; Angelidaki, I.; Kalyuzhnyi, S.V.; Pavlostathis, S.G.; Rozzi, A.; Sanders, W.; Siegrist, H.; Vavilin, V. *Anaerobic Digestion Model No. 1 (ADM1), Report no. 13*; International Water Association: London, UK, 2002.
74. Mata-Alvarez, J. (Ed.) Fundamentals of the Anaerobic Digestion Process. In *Biomethanization of the Organic Fraction of Municipal Solid Wastes*; International Water Associate: London, UK, 2005; Volume 4, pp. 1–20. ISBN 978-1-78040-299-4.
75. Amodeo, C.; Hafner, S.D.; Franco, R.T.; Benbelkacem, H.; Moretti, P.; Bayard, R.; Buffière, P. How different are manometric, gravimetric and automated volumetric BMP results? *Water* **2020**. in preparation.
76. Holliger, C.; Fruteau de Laclos, H.; Hafner, S.D.; Koch, K.; Weinrich, S.; Astals, S.; Alves, M.; Andrade, D.; Angelidaki, I.; Appels, L.; et al. Requirements for Measurement of Biochemical Methane Potential (BMP). Standard BMP Methods Document 100, Version 1.3. Available online: <https://www.dbfz.de/en/BMP> (accessed on 19 April 2020).
77. Shah, T.A.; Ullah, R. Pretreatment of wheat straw with ligninolytic fungi for increased biogas productivity. *Int. J. Environ. Sci. Technol.* **2019**, *16*, 7497–7508. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).