Cleaner Chemical Engineering 4 (2022) 100065

Contents lists available at ScienceDirect

Cleaner Chemical Engineering

journal homepage: www.elsevier.com/locate/clce

# Assessing the value of kinetic results from biochemical methane potential tests: Reproducibility from a large inter-laboratory study

Sasha D. Hafner[a,*], Sergi Astals[b], Christof Holliger[c], Konrad Koch[d,1], Lisa Nielsen[a], Lina Refsahl[a], Sören Weinrich[e]

[a] Department of Biological and Chemical Engineering, Aarhus University, Gustav Wiedsvej 10C, Aarhus, 8000, DK
[b] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franqués 1, Barcelona, 08028, Spain
[c] Ecole Polytechnique Fédérale de Lausanne, ENAC-IIE Laboratory for Environmental Biotechnology, Station 6, 1015 Lausanne, Switzerland
[d] Chair of Urban Water Systems Engineering, Technical University of Munich, Am Coulombwall 3, Garching, 85748, Germany
[e] Biochemical Conversion Department, Deutsches Biomasseforschungszentrum gemeinnützige GmbH, Torgauer Straße 116, Leipzig, 04315, Germany

ARTICLE INFO

*Keywords:*
Anaerobic digestion
Hydrolysis rate
Biomethane potential (BMP)
First-order kinetics
Parameter estimation

ABSTRACT

Kinetic information extracted from biochemical methane potential (BMP) tests is often reported but its value is unclear. Inter-laboratory reproducibility provides a useful indication of its value. Here we extracted estimates of the first-order rate constant $k$ from 1259 methane production curves collected in a large inter-laboratory study on BMP in order to quantify reproducibility. Reproducibility in $k$ was poor; relative standard deviation was 50–140%. Substrate comparisons ($k$ for one substrate compared to another) also had low reproducibility, regardless of low $p$ values from inferential statistical tests. The use of a shared inoculum did not improve reproducibility in $k$. We conclude that $k$ estimates from BMP tests only partially reflect intrinsic substrate properties. Therefore, interpretation and application of batch kinetic results should be done cautiously.

## 1. Introduction

Batch biochemical methane potential (BMP) tests generate information on the rate of methane production in addition to the potential quantity that is available. This information is typically extracted by application of a mathematical model, and reported as a rate constant, e.g., $k$ ($d^{-1}$) for a first-order model, and an "ultimate" potential, often called $B_0$ (mL $g^{-1}$). Despite some discussion on limitations of these values (Brulé et al., 2014; Da Silva et al., 2022; Donoso-Bravo et al., 2019; Guo et al., 2021; Koch et al., 2020), they are commonly reported in the literature, and occasionally applied. There are at least two common applications: 1) to full-scale reactors or other systems for process design, efficiency evaluation, monitoring, or control (García-Gen et al., 2015; Insel et al., 2022), and 2) to compare or rank substrates or evaluate pre-treatments (El Gnaoui et al., 2022; Li et al., 2017).

The focus of the present work is the value of the rate constant. Although both $k$ and $B_0$ influence any prediction of $CH_4$ production rate, $k$ is more difficult to precisely determine. In contrast, even the measured BMP (e.g., $BMP_{1\%,3d}$) provides a reasonable estimate of ultimate BMP without any model application (see Section 3.3). Although reproducibility in BMP is a problem, the magnitude of variability is much smaller than in $k$, and procedures have been developed for improving reproducibility (Hafner et al., 2020).

Few studies have compared batch and full-scale (or even lab-scale batch and lab-scale continuous) kinetic estimates. Comparisons have shown similar estimates of first-order rates (Chynoweth et al., 1993), slightly lower rates (Jensen et al., 2009) or much lower rates (Batstone et al., 2009). In general, these limited results suggest that reliability of batch kinetic measurements to describe full-scale processes is low.

The second application listed above may be assumed to be more reliable, but reproducibility in these types of comparisons has not been previously evaluated in a large inter-laboratory study. An increase in conversion rate is a typical objective of pre-treatment (Carrere et al., 2016), but exactly how reproducible rate effects are is not clear.

Accurate assessment of kinetic constants is important; assuming substrate conversion follows first-order kinetics, a 10-fold change in $k$ requires a proportional change in retention time to maintain conversion efficiency in a continuous reactor. In general, time to a fixed conversion efficiency or conversion at a fixed retention time is sensitive to rate constant magnitude.

* Corresponding author.
*E-mail addresses:* sasha.hafner@au.dk (S.D. Hafner), sastals@ub.edu (S. Astals), k.koch@tum.de (K. Koch), soeren.weinrich@dbfz.de (S. Weinrich).
[1] URL:www.cee.ed.tum.de/sww

https://doi.org/10.1016/j.clce.2022.100065
Received 22 June 2022; Received in revised form 27 August 2022; Accepted 15 September 2022

Some degree of reproducibility is required if batch rate constant estimates are to have any value. Reproducible differences between substrates implies that batch results reflect intrinsic substrate characteristics, but not that estimated rate constants are necessarily the same in full-scale systems. Conversely, very poor reproducibility in batch kinetics precludes any possibility of batch results consistently reflecting full-scale behavior or even intrinsic differences between substrates. Therefore, an evaluation of reproducibility in $k$ determined in batch tests can serve to assess the quality of these commonly reported values.

The goal of this work was to assess the quality of kinetic information from BMP tests, by quantifying reproducibility in rate constant estimates. To do this, the first-order rate constant $k$ was extracted from more than 1200 specific methane production (SMP) curves of four substrates collected in a large inter-laboratory study. These estimates were then used to quantify inter-laboratory reproducibility and assess some prospective approaches for improving it.

## 2. Methods

### 2.1. Experimental data

1259 SMP curves (cumulative $CH_4$ yield and average interval production rate) were taken from a large inter-laboratory study (Hafner et al., 2020). Data are available at https://github.com/sashahafner/BMP-kinetics-paper-2022 and data collection is described in (Hafner et al., 2020). Measurements were made in laboratories in 14 different countries, almost exclusively from Europe (see the acknowledgments section in Hafner et al., 2020). Three sets of BMP tests were carried out: two in a first period referred to as "study S1" and one in "study S2". SMP values were calculated from raw data as net $CH_4$ production from substrate per g added substrate volatile solids (VS). Substrates were dried and finely ground, and varied in composition and degradability (rate and extent). Three substrates were mixed animal feed components (SA, SB, and SC), one was wheat straw (SD), and microcrystalline cellulose (CEL) was included as a positive control with known chemical composition. Laboratories followed a general protocol based on Holliger et al. (2016), with inoculum-to-substrate ratio of 2:1 (or higher in some cases) and mesophilic incubation (35–40 °C), but did not use identical methods for measuring biogas production and composition. The most common measurement methods were the AMPTS II system (BPC Instruments AB, Sweden) along with manual volumetric and manometric methods. A subset of laboratories in France, Germany, and Italy carried out tests with two inocula: from their typical source and a single shared source within each country. For more details see Hafner et al. (2020).

### 2.2. First-order model

Extraction of kinetic information generally requires fitting a model. Here, a first-order rate constant and ultimate BMP were extracted from each SMP curve using a simple approach: a first-order model applied after exclusion of any lag phase. Several factors were considered in selecting the approach used for model fitting. Inspection of the measured $CH_4$ production rates showed that rates generally declined over time, consistent with a first-order model. But an early period of low and slowly increasing production rate was common, although highly variable in duration and $CH_4$ production rate. We assumed *a priori* that quantitative characteristics of this "lag phase" contains minimal useful information about intrinsic substrate degradation rate (Koch et al., 2019), as it is likely a product of the starving inoculum being exposed to a high substrate availability at the beginning of the batch test. Instead of trying to capture this complex behavior that may be unique for each individual bottle, we excluded it and focused on the period during which $CH_4$ production appears to be limited by substrate quantity. Note that our use of the term "lag phase" in this work (and other biogas applications) is different from the definition used in microbiology, where it refers to a period of no net increase in cell numbers (Yates and Smotzer, 2007). By design, BMP tests have a large quantity of active microorganisms relative to the available substrate, to ensure that degradation is relatively rapid and complete. A first-order response is common in BMP tests, and it has been proposed that significant deviation from this response may indicate a problem in the test (Koch et al., 2019).

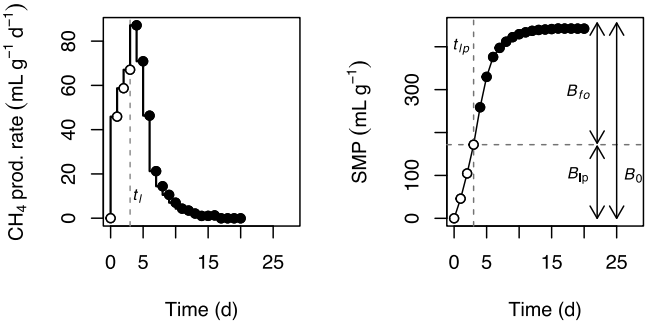


**Fig. 1.** A graphical representation of the approach used for kinetic parameter extraction (Section 2.2) based on a single SMP curve. Open circles show measurements within the "lag phase".

In the approach used here, SMP or cumulative $CH_4$ yield is given by:

$$y(t) = B_{fo}(1 - e^{-k(t-t_{lp})}) + B_{lp} \quad (1)$$

where $y(t)$ = specific methane production (SMP) (mL $g^{-1}$, net standardized $CH_4$ volume per g substrate VS), $B_{fo}$ = ultimate SMP in the first-order period after any lag phase (mL $g^{-1}$), $k$ = first-order kinetic constant ($d^{-1}$), $t$ = elapsed time of BMP test (d), $t_{lp}$ = duration of lag phase (d), and $B_{lp}$ = $CH_4$ production during lag phase (i.e. SMP at $t_l$) (mL $g^{-1}$). SMP at infinite time or "ultimate" BMP $B_0 = B_{fo} + B_{lp}$. With Eq. (1), the lag phase can be ignored (not excluded) by setting $t_{lp}$ and $B_{lp}$ to zero.

In practice, time $t$ is discrete; measurements are made at particular times at the end of each incubation interval. What is typically measured is actually the average rate of change in cumulative SMP:

$$\bar{r}_i = \frac{\Delta y}{\Delta t} = \frac{y_{t_i} - y_{t_{i-1}}}{t_i - t_{i-1}} \quad (2)$$

This rate can be calculated from measured or model-calculated (Eq. (1)) SMP ($y$) using Eq (2). Or, substitution of Eq. (1) into Eq. (2) gives a direct alternative for model calculations:

$$\bar{r}_i = \frac{B_{fo}e^{kt_{lp}}(e^{-kt_{i-1}} - e^{-kt_i})}{t_i - t_{i-1}} \quad (3)$$

where $\bar{r}_i$ = average methane production rate (mL $d^{-1}$ $g^{-1}$, net standardized $CH_4$ volume per g substrate VS over the period $t_{i-1}$ to $t_i$). With long incubation intervals (difference of several days between $t_i$ and $t_{i-1}$) this average rate may be substantially different from the instantaneous rate, which is given by the derivative of Eq. (1), because the model is nonlinear with respect to time (rate vs. time).

The parameters $t_{lp}$ and $B_{lp}$ are determined directly from measurements prior to model application (Fig. 1). In this study, the lag phase is defined as the period when $t <$ time of maximum average rate, i.e., $t_{lp}$ = the end of the interval prior to the interval with the maximum measured rate ($\bar{r}_i$).

This definition derives from the behavior of a first-order model (rate must always decrease over time), but also reflects the period in which $CH_4$ production may be limited by the microbial community and not substrate quantity, and therefore not reflect any intrinsic rate of substrate degradation.

### 2.3. Data analysis

#### 2.3.1. Parameter estimation

Least-squares estimates of the two model parameters $k$ and $B_{fo}$ were determined for each separate SMP curve through nonlinear regression,

using the `fitFOM()` function in the biogas package (Hafner et al., 2018) (v1.40, https://github.com/sashahafner/biogas/releases/tag/v1.40), which uses the Levenberg-Marquardt optimization algorithm implemented in the R package minpack.lm (Elzhov et al., 2016). The response variable was $\bar{r}$, calculated separately for each measurement interval for each individual bottle, and the objective function was the sum of squares. Parameter values were log-transformed during estimation to avoid negative values.

For curves that showed graphical evidence of two phases in $CH_4$ production, Eq. (2) was extended to include two kinetic fractions of slowly and rapidly degradable substrate components, i.e., a two-pool model, to demonstrate limitations of the alternative. This approach was similar to "Model C" described in Brulé et al. (2014). Variations in parameter extraction (combinations that included fitting to cumulative SMP, omitting exclusion of lag phase observations, and forcing the model through the last SMP value, for a total of six approaches) were explored as well.

#### 2.3.2. Analysis of rate constants

The primary response variable was first-order rate constant $k$, which was $\log_{10}$-transformed to deal with high variability, approximately log-normal distributions, and correlation between variability and magnitude. This transformation makes all reported comparisons in $k$ magnitude relative. Median parameter values were calculated for each set of 3 (occasionally 2) bottles for a single study (S1 and S2) × test (T1 and T2) × laboratory × substrate combination, and these values were used as observations in the following analysis. Precision among replicates, a measure of repeatability, was quantified as standard deviation among these values.

Inter-laboratory reproducibility in extracted $k$ was quantified using standard deviation among laboratories, applied separately for each study × test × substrate combination (as Hafner et al. (2020) did for BMP). To group estimated rate constants $k$ extreme "ultimate" BMP values $B_0$ were identified as those below 50% or above 120% of theoretical maximum BMP. Theoretical maximum BMP was calculated from elemental composition assuming complete conversion and no biomass production using Eq. 13.6 in (Rittmann and McCarty, 2001) as implemented in the `predBg()` function in the biogas package in R (Hafner et al., 2018). No estimates of uncertainty in individual parameter values for individual bottles were made because interpretation is ambiguous for nonlinear regression in general (Motulsky and Ransnas, 1987), and questions about both independence of observations and the accuracy of the model structure render them even less useful. Furthermore, such estimates do not clearly add value beyond the assessment of repeatability and reproducibility carried out.

The ratios of rate constants (each substrate compared to CEL, i.e., normalized rate constants) were also used to assess the reproducibility of relative differences and determine if normalization might improve reproducibility (Donoso-Bravo et al., 2019). The relationships between reproducibility and BMP validation criteria and measurement method was assessed graphically or by calculating standard deviation for subsets as above. Boxplots were used to show variation within and among substrates; they show the median, 25th and 75th percentiles (the box), minimum and maximum (whiskers), and outliers (points, defined here as values more than 1.5 time the interquartile range beyond the edge of the range). Finally, the importance of substrate × laboratory interactions was explored by simple separate (for each BMP test or lab) hypothesis tests comparing CEL and SC, CEL and SD $k$ using $t$-tests, as well as analysis of variance (ANOVA) with test (lab) × substrate as predictor variables.

## 3. Results and discussion

### 3.1. Qualitative description of SMP curves

Most SMP curves showed a similar pattern: an increase in average $CH_4$ production rate over a few days (defined as the lag phase), followed by a continual decline through the remainder of the incubation. Figure 2 shows representative curves for six individual bottles, and curves for all individual bottles can be found in the data repository associated with this work (https://github.com/sashahafner/BMP-kinetics-paper-2022). The presence of a lag phase was ubiquitous; only for SA and SB did the frequency of a clear lag phase drop below 95% (70–90% for SA, 60% for SB). Lag phases were not typically long. Median values were 3 d or lower for all substrates, and highest for cellulose. However, these short periods still accounted for a substantial fraction of total $CH_4$ production: 25–30% of final cumulative SMP for SC, and less for other substrates. Directly addressing the lag phase problem by excluding it improved model fit and reduced inter-laboratory variability for both $k$ and $B_0$.

Model efficiency (ME) (Nash and Sutcliffe, 1970) was generally high, suggesting that measurements approximately followed a single-pool first-order model (excluding any lag phase). Median ME was > 0.90 for all substrates. Differences in ME reflect qualitative differences in the shape of SMP curve. Some curves clearly deviated from a single-pool first-order model even after exclusion of the lag phase, as shown by part C of Fig. 2. This particular example, as well as some others, followed a two-pool model. Bottle 1012 showed a response completely different from a first-order model, and is an example of a curve that should not be used for parameter extraction, as shown in part F of Fig. 2. Visual examination of this type of plot (Fig. 2) can show a mismatch between measurements and model structure as well as other types of problems, and is an important step in assessing the quality of parameter estimates (Motulsky and Ransnas, 1987). The scale of the present work makes the task implausible here, but plots of rate and SMP curves from the data repository associated with this work (https://github.com/sashahafner/BMP-kinetics-paper-2022) do show clear examples where parameter estimates should not be used, including bottles 160, 548, 1009, 1162, and 1681.

### 3.2. Parameter value summary

Estimates of $k$ and $B_0$ were weakly negatively correlated (Fig. 3), which might be expected due to compensation between parameter estimates for $k$ and $B_0$. Correlation reflects some of the uncertainty in determining model parameters from indirect measurements, and implies that implausible $B_0$ estimates can signal a problem with $k$ estimates or the overall model structure. In some cases, extreme $k$ values were associated with extreme ultimate BMP, and these values could readily be flagged as inaccurate. Model ultimate BMP estimates were close to measured 1% net 3 d BMP values in almost all cases, but were often slightly lower (median values ranged from 0.96 to 1.04 for all substrate × text combinations). In many cases this difference was related to underestimation of late low production rates (e.g., Fig. 2, case C). Although the combination of parameter estimates and the single-pool first-order model (Section 2.2) would tend to underestimate ultimate BMP, this mismatch between reality and the model is not necessarily a major problem for extraction of kinetic information. But this issue should serve as a reminder that batch kinetics often do not strictly follow first-order behavior, and that SMP curves from different laboratories can differ qualitatively, even for the same substrate.

### 3.3. Rate constant precision

Inter-laboratory reproducibility in $k$ was generally low (high variability) and somewhat substrate-dependent. Relative standard deviation among labs was 50 to 140% (Table 1). Differences between extremes were > 10-fold for most substrates. Substrate SD, wheat straw, which had the slowest degradation rate, had the lowest variability. The positive control CEL was not different from complex substrates.

In contrast to reproducibility among different labs, repeatability within each lab, or precision of individual measurements ($n$ = 3 or in some cases, 2) was generally high: median relative standard deviation

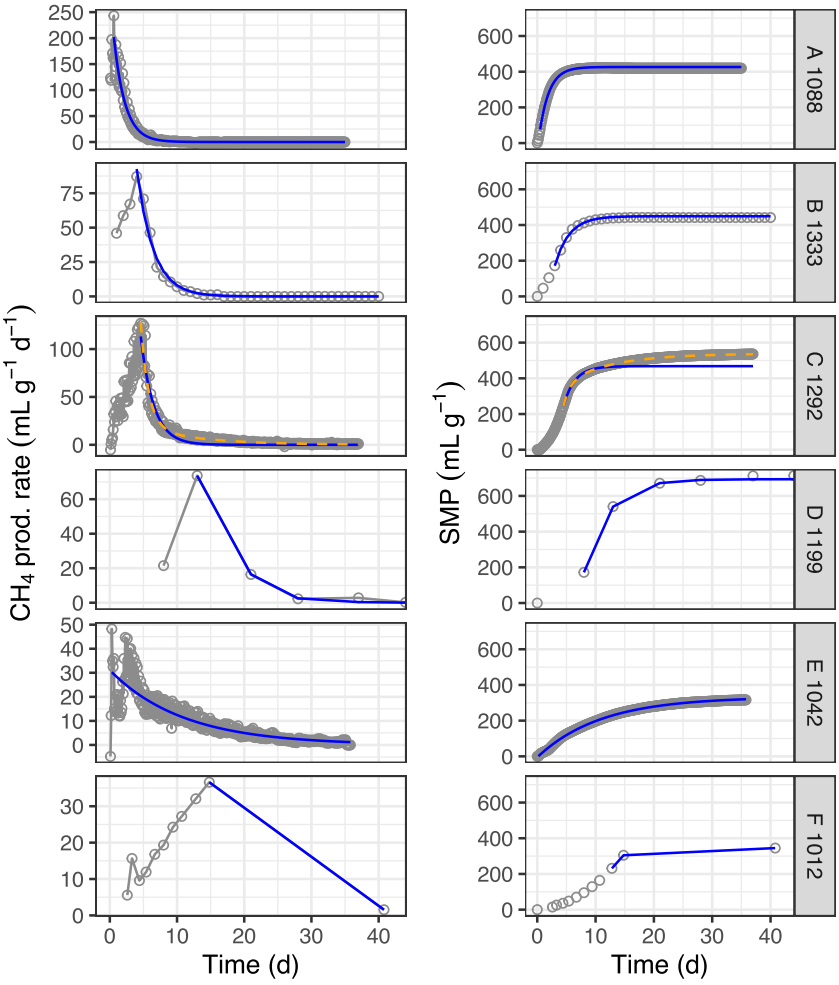


**Fig. 2.** Examples of methane production curves from six bottles, showing average rates (left column) and cumulative specific methane production (right column). Plots display both measured values (gray points and lines) and model calculations (solid blue: single fraction, dashed orange: 2-pool model (only in part C)). For rates (left), the position of plotted values on x axis is the end of each sampling interval, and average rate is shown on y axis. Selected examples show: A, short lag phase and first-order response; B, long lag phase and first-order response; C, 2-pool first-order response (and long lag pase); D, low resolution measurements; E, no clear lag phase and low degradation rate; F, low but increasing rate, not a first-order response. Integer plot labels show bottle ID code present in the measurement data. All curves are for substrate SC except E (SD) and F (CEL). The plots in the data repository (see data availability section) show even more diversity. Note that the y axis scales vary for the left but not right plots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

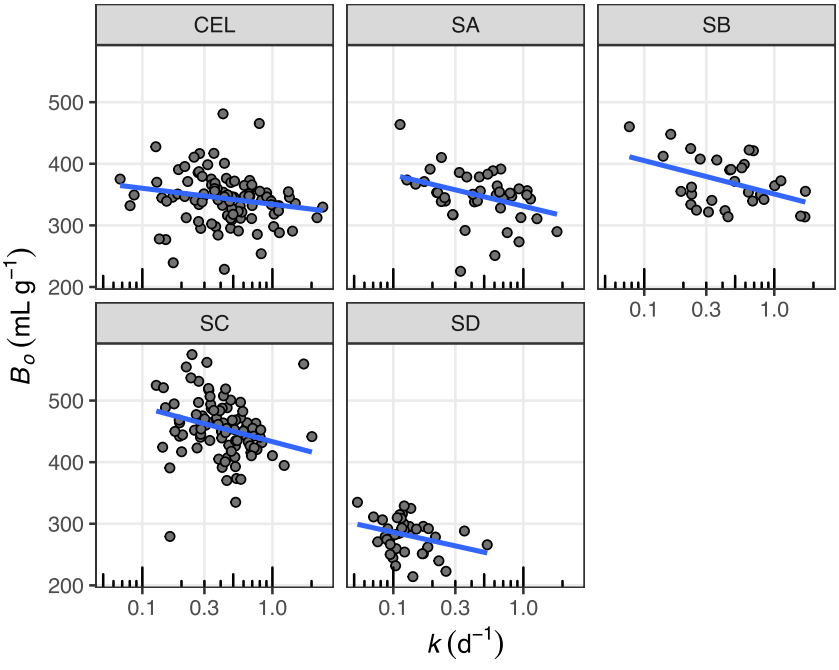


**Fig. 3.** Best-fit values of ultimate $CH_4$ yield $B_0$ vs. rate constant $k$ by substrate. Observations with extreme values of $B_0$ (Section 2.3.2) were excluded from plots. Lines show robust regression results.

**Table 1**
Summary of best-fit $k$ values for all laboratories by test and study.

| | | | | Median | | | Std. dev. | | Range |
|---|---|---|---|---|---|---|---|---|---|
| Substrate | Study | Test | $n$ | $t_{lp}$ (d) | $k$ ($d^{-1}$) | $B_0$ (mL $g^{-1}$) | $k$ (%) | $B_0$ (%) | $k$ (-fold) |
| CEL | S1 | T1 | 21 | 2.0 | 0.42 | 345 | 88 | 13 | 10 |
| CEL | S1 | T2 | 21 | 2.5 | 0.42 | 348 | 89 | 15 | 13 |
| CEL | S2 | T1 | 66 | 3.0 | 0.48 | 339 | 111 | 24 | 30 |
| SA | S1 | T1 | 21 | 0.7 | 0.46 | 352 | 113 | 14 | 14 |
| SA | S1 | T2 | 21 | 1.0 | 0.53 | 349 | 83 | 14 | 9 |
| SB | S1 | T1 | 19 | 0.5 | 0.46 | 355 | 142 | 19 | 33 |
| SB | S1 | T2 | 19 | 0.0 | 0.48 | 342 | 115 | 32 | 16 |
| SC | S1 | T1 | 20 | 1.0 | 0.43 | 458 | 79 | 22 | 9 |
| SC | S1 | T2 | 20 | 1.5 | 0.43 | 454 | 57 | 16 | 5 |
| SC | S2 | T1 | 60 | 2.0 | 0.41 | 453 | 96 | 20 | 47 |
| SD | S2 | T1 | 43 | 2.0 | 0.12 | 284 | 52 | 19 | 10 |

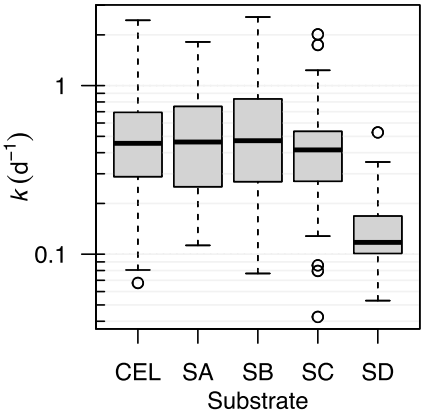


**Fig. 4.** Inter-laboratory variability in best-fit first-order rate constant $k$ by substrate. Note the logarithmic scale (y axis); the interquartile range is 2- to 4-fold for most substrates.

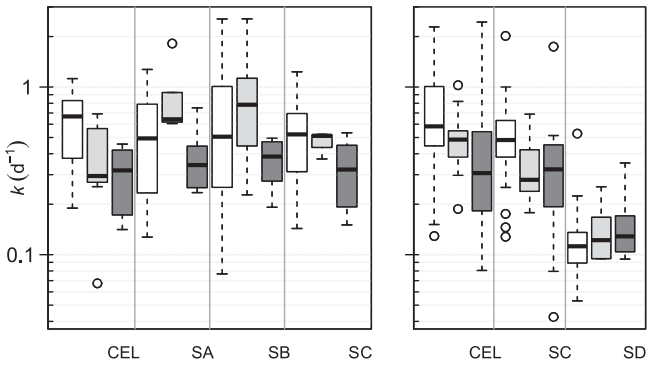


**Fig. 5.** Best-fit rate constant $k$ plotted by substrate, study (S1, left, or S2, right), and measurement method (white = AMPTS II, light gray = manometric, dark gray = other volumetric).

among replicates was only 6–18% depending on substrate (lowest for SA).

Normalization to CEL $k$ slightly improved variability for most substrates (standard deviation decreased by 0 to 46%, excluding SD and a single SC combination) (Fig. 4). But for the slowest degrading substrate SD, variability in normalized $k$ was even higher (standard deviation increased by a factor of 3), which is not surprising considering that variability in $k$ was lower for SD than CEL. This difference for SD may suggest that degradation rate of substrates with higher availability of easily degradable components (SA, SB, SC were animal feeds after all) is more sensitive to inoculum properties, but this hypothesis needs testing.

As the normalization results hint, apparent laboratory × substrate interactions contributed to low reproducibility in intra-laboratory substrate comparisons. This was shown by an $F$ test result using all substrates ($p < 2 \cdot 10^{-16}$ for an interaction term) and comparisons made using individual $t$-tests applied by laboratory (test) to CEL and SC only. Nearly half of $t$-tests (56/118) had $p < 0.01$ but of these, those with SC $k$ > CEL $k$ (24) were nearly as common as those with the opposite result (32). This is an important result: Not only is it difficult to compare substrates analyzed in different laboratories, it cannot be assumed that a difference between substrates observed in one laboratory is representative of the general response, regardless of statistical significance (the lowest observed $p$ value was below $1 \cdot 10^{-7}$, and relative differences in mean $k$ ranged from 0.13× to 4.1×). However, an overall large difference in mean $k$ coupled with a low $p$ value may still be meaningful, as shown by a similar comparison between SD and CEL. Here, all tests with $p < 0.01$ (48/59) showed lower $k$ for SD.

Application of current BMP validation criteria did not clearly improve reproducibility, but did have some effects. Best-fit $k$ was higher for BMP-validated results for at least some substrates (+24% for overall effect of validation, $p = 0.002$). Most lack of validation was caused by a low cellulose BMP. For some substrates, validation criteria were able to eliminate extreme $k$ values (2 of 3 for SC and 7 of 11 for SD eliminated) and improved reproducibility for SC, and SD (17 to 60% reduction).

There were no consistent differences in variability among measurement methods (Fig. 5). All three popular methods were variable for some substrate × study combinations, and showed similar results for SD (gravimetric and absolute GC were excluded because they were used by few labs and variability was low in most cases). In contrast, there were differences in $k$ magnitude: other volumetric methods provided lower $k$ values than AMPTS II in both studies S1 (-52%) and S2 (-23%) ($p < 0.03$ by $F$-test) (Fig. 5). Mechanical mixing and a high measurement frequency for the commercial AMPTS II system, compared to other (typically manual) volumetric methods both may have played a role.

Use of a shared inoculum did not consistently reduce variability in measurement of $k$ (Fig. 6). Variability increased in as many cases as it decreased (4 of 8) (standard deviation calculated with mean values for each laboratory × substrate combination, i.e., points in Fig. 6). Even with an identical inoculum source, different laboratories measured different rates (even exceeding a factor of 3) and curves showed different development of rate over time. Why inoculum sharing was not sufficient is not clear, but disruption due to transport, transfer, and storage is a possible source of error. An analysis of BMP measurements from this same dataset showed no evidence of a general improvement in the reproducibility of BMP by sharing inoculum either (Hafner et al., 2020).

In general, results were not dependent on the exact approach used for extracting parameters. Fitting through the lag phase tended to result

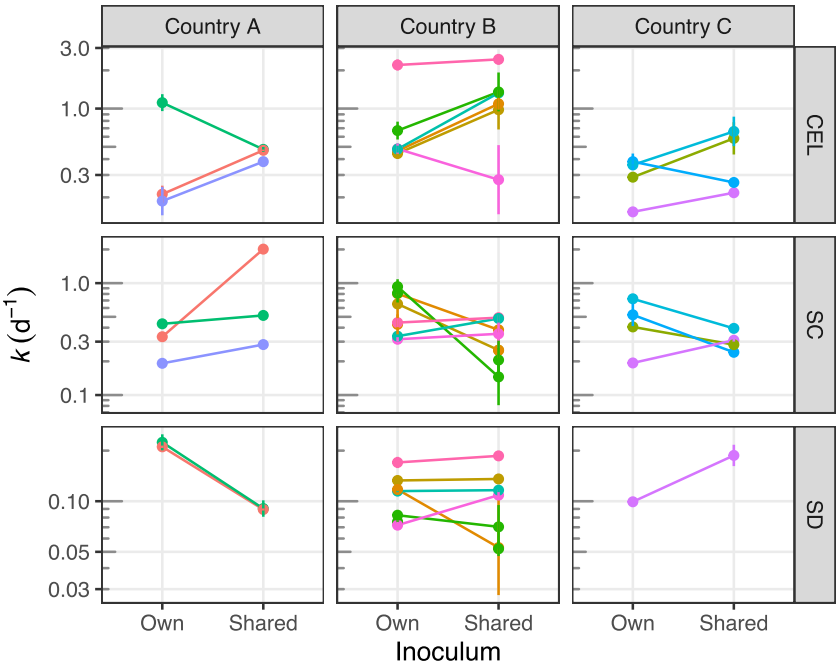


**Fig. 6.** Best-fit $k$ values versus inoculum source. Lines connect points from the same laboratory and each individual laboratory is represented by a unique color. Inoculum was shared within countries. Points show mean values (typically $n = 3$) and vertical lines show standard deviation.

in lower $k$ estimates and more variability in $k$ and $B_0$. Using cumulative SMP ($y$ instead of rate $\bar{r}$) as the response variable decreased variability, but these observations are not independent and by definition the approach presented here more accurately represents measured rates (Sections 2.2 and 2.3.1). Details for a total of six different approaches applied can be found in the data repository that is associated with this paper (https://github.com/sashahafner/BMP-kinetics-paper-2022). Furthermore, because the SMP measurements used in this work are now publicly available through this repository, others are free to repeat this analysis using different approaches.

### 3.4. Can reproducibility be improved?

The magnitude of $k$ was positively correlated with the quantity of $CH_4$ produced during the lag phase (Spearman's rank correlation coefficient $\rho = 0.4$) and weakly correlated with ME ($\rho = 0.2$). However, there was no clear relationship between reproducibility and these characteristics, i.e., neither shows strong potential for development of validation criteria. The presence of consistent correlation undermines any assumption that $k$ reflects only intrinsic properties of the substrate.

Model fit and the ratio of $B_0$ to measured BMP (1% net 3 d) were related. For very high ME, $B_0$ must be larger than measured BMP, but typically only slightly larger. The value of $k$ was strongly negatively correlated with the ratio $B_0$:BMP ($\rho = -0.66$), which probably reflects true differences in kinetics (not necessarily intrinsic). Values below unity indicate that the single-fraction first-order model used does not completely describe the observed response, and at least some extreme $k$ estimates could be eliminated by applying a $B_0$:BMP cutoff. However, it is not reasonable to exclude all results with a ratio below 1.0, considering that all parameter estimates will have error, and the estimates of $k$ do not qualitatively change below this value.

Rate constants extracted from SMP curves clearly reflect more than intrinsic characteristics of substrates, probably including effects of inocula and other differences related to the BMP test protocol that are evidently difficult to standardize (even with the use of the same inoculum source). Although there has recently been significant success in improving BMP reproducibility through application of validation criteria, drastic improvement in the reproducibility of $k$ seems unlikely given these results. Kinetic results may simply be too sensitive to the test environment or the initial state and development of the microbial community, which in turn may respond to small differences in handling and other test procedures. Stated differently, results suggest that $k$ values from batch tests depend on numerous factors, and intrinsic substrate characteristics is only one.

Although no indicators show potential for a drastic improvement in reproducibility as validation criteria, the most reproducible $k$ estimates likely come from cases where the applied first-order model accurately describes the measurements, i.e., cases which have high ME and a ratio of $B_0$ (model ultimate BMP) to measured BMP near but slightly greater than 1. Furthermore, low or no lag phase $CH_4$ production and a high sampling resolution (possibly quantified by a small maximum value of $\bar{r}$) indicate that extracted values are based on a high number of measurements, and these might also serve as validation criteria. However, defining quantitative validation criteria for $k$ is arbitrary, and selecting values that will be widely useful and not too restrictive is challenging and of dubious value at this time, at least for the present data set. With few exceptions, changes in $k$ with respect to potential indicators are gradual. Further complicating this task is the issue with variability among substrate types. How representative are the five substrates used here? Substrates varied in composition and degradability (Section 2.1) but were all dried, finely ground, and primarily carbohydrate-based. While it is likely that other substrates may show different results, there is no clear reason to expect that the low inter-laboratory reproducibility quantified here is not generally representative. It is possible that less homogeneous or more complex materials may show even more variability. Regardless, the minimal step of applying current BMP validation criteria (Holliger et al., 2021) is recommended. A visual comparison of the curve calculated by the model and measurements should also always be carried out (Motulsky and Ransnas, 1987).

## 4. Conclusions

Inter-laboratory reproducibility in first-order rate constants from batch tests was poor and not improved by sharing an inoculum. BMP validation criteria slightly improved rate constant reproducibility and is recommended, but normalization to a reference substrate did not consistently improve reproducibility and cannot be recommended. Results show that rate constants from BMP tests should be taken only as approximate indicators of the substrate degradation rate and any application of these types of results should account for associated uncertainty.

Low reproducibility also implies that substrate and treatment comparisons should be done in a single batch test within a single laboratory. Nonetheless, due to laboratory effects that vary among substrates, differences in rate constants may not represent a general difference that would be observed in other laboratories, highlighting the importance of repetition.

## Supplementary data

E-supplementary data of this work can be found in online version of the paper. A single pdf file with a table and plots showing data or comparisons that support or clarify some results described in the paper is available.

## Acknowledgments

This work would not have been possible without the generous contributions of time and expertise by the more than 50 individuals from nearly 40 laboratories who carried out the BMP tests as part of the IIS-BMP project. The Swiss Federal Office for Energy funded the management part of the IIS-BMP inter-laboratory project that generated the data used here, as well as a portion of the work on kinetics. Funding from the TUM Global Incentive Fund to host the second international workshop on standardization of BMP tests in Freising (Germany) is highly appreciated. Sergi Astals is grateful to the Spanish Ministry of Science, Innovation and Universities for his Ramon y Cajal fellowship (RYC-2017-22372). Sören Weinrich is thankful for funding from the German Federal Ministry of Food and Agriculture to the junior research group on simulation, monitoring and control of anaerobic digestion plants (grant 2219NR333).

## Data availability

Data are available from a public repository: https://github.com/sashahafner/BMP-kinetics-paper-2022. The repository includes details on BMP tests, specific methane production (SMP), and kinetic results both summarized and for individual bottles. Plots show SMP and rate curves for all individual bottles, and include model results. A detailed description of the contents can be found in the README.md file.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Sasha D. Hafner:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft. **Sergi Astals:** Conceptualization, Methodology, Writing – review & editing. **Christof Holliger:** Writing – review & editing, Project administration, Funding acquisition. **Konrad Koch:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition. **Lisa Nielsen:** Methodology, Investigation, Writing – review & editing. **Lina Refsahl:** Methodology, Investigation, Writing – review & editing. **Sören Weinrich:** Conceptualization, Methodology, Writing – review & editing.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.clce.2022.100065.

## References

Batstone, D.J., Tait, S., Starrenburg, D., 2009. Estimation of hydrolysis parameters in full-scale anerobic digesters. Biotechnol. Bioeng. 102 (5), 1513–1520. doi:10.1002/bit.22163.

Brulé, M., Oechsner, H., Jungbluth, T., 2014. Exponential model describing methane production kinetics in batch anaerobic digestion: a tool for evaluation of biochemical methane potential assays. Bioprocess Biosyst. Eng. 37 (9), 1759–1770. doi:10.1007/s00449-014-1150-4.

Carrere, H., Antonopoulou, G., Affes, R., Passos, F., Battimelli, A., Lyberatos, G., Ferrer, I., 2016. Review of feedstock pretreatment strategies for improved anaerobic digestion: from lab-scale research to full-scale application. Bioresour. Technol. 199, 386–397. doi:10.1016/j.biortech.2015.09.007.

Chynoweth, D., Turick, C., Owens, J., Jerger, D., Peck, M., 1993. Biochemical methane potential of biomass and waste feedstocks. Biomass Bioenergy 5 (1), 95–111. doi:10.1016/0961-9534(93)90010-2.

Da Silva, C., Peces, M., Faundez, M., Hansen, H., Campos, J.L., Dosta, J., Astals, S., 2022. Gamma distribution function to understand anaerobic digestion kinetics: kinetic constants are not constant. Chemosphere 306, 135579. doi:10.1016/j.chemosphere.2022.135579.

Donoso-Bravo, A., Ortega, V., Lesty, Y., Bossche, H.V., Olivares, D., 2019. Addressing the synergy determination in anaerobic co-digestion and the inoculum activity impact on BMP test. Water Sci. Technol. 80 (2), 387–396. doi:10.2166/wst.2019.292.

El Gnaoui, Y., Frimane, A., Lahboubi, N., Herrmann, C., Barz, M., EL Bari, H., 2022. Biological pre-hydrolysis and thermal pretreatment applied for anaerobic digestion improvement: kinetic study and statistical variable selection. Cleaner Waste Syst. 100005. doi:10.1016/j.clwas.2022.100005.

Elzhov, T. V., Mullen, K. M., Spiess, A.-N., Bolker, B., 2016. Minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds. https://CRAN.R-project.org/package=minpack.lm.

García-Gen, S., Sousbie, P., Rangaraj, G., Lema, J.M., Rodríguez, J., Steyer, J.-P., Torrijos, M., 2015. Kinetic modelling of anaerobic hydrolysis of solid wastes, including disintegration processes. Waste Manage. 35, 96–104. doi:10.1016/j.wasman.2014.10.012.

Guo, H., Oosterkamp, M.J., Tonin, F., Hendriks, A., Nair, R., van Lier, J.B., de Kreuk, M., 2021. Reconsidering hydrolysis kinetics for anaerobic digestion of waste activated sludge applying cascade reactors with ultra-short residence times. Water Res. 202, 117398. doi:10.1016/j.watres.2021.117398.

Hafner, S.D., Fruteau de Laclos, H., Koch, K., Holliger, C., 2020. Improving inter-laboratory reproducibility in measurement of biochemical methane potential (BMP). Water 12 (6), 1752. doi:10.3390/w12061752.

Hafner, S.D., Koch, K., Carrere, H., Astals, S., Weinrich, S., Rennuit, C., 2018. Software for biogas research: tools for measurement and prediction of methane production. SoftwareX 7, 205–210. doi:10.1016/j.softx.2018.06.005.

Holliger, C., Alves, M., Andrade, D., Angelidaki, I., Astals, S., Baier, U., Bougrier, C., Buffière, P., Carballa, M., de Wilde, V., Ebertseder, F., Fernández, B., Ficara, E., Fotidis, I., Frigon, J.-C., Fruteau de Laclos, H., S. M. Ghasimi, D., Hack, G., Hartel, M., Heerenklage, J., Sarvari Horvath, I., Jenicek, P., Koch, K., Krautwald, J., Lizasoain, J., Liu, J., Mosberger, L., Nistor, M., Oechsner, H., Oliveira, J.V., Paterson, M., Pauss, A., Pommier, S., Porqueddu, I., Raposo, F., Ribeiro, T., Rüsch Pfund, F., Strömberg, S., Torrijos, M., van Eekert, M., van Lier, J., Wedwitschka, H., Wierinck, I., 2016. Towards a standardization of biomethane potential tests. Water Sci. Technol. 74 (11), 2515–2522. doi:10.2166/wst.2016.336.

Holliger, C., Astals, S., de Laclos, H.F., Hafner, S.D., Koch, K., Weinrich, S., 2021. Towards a standardization of biomethane potential tests: a commentary. Water Sci. Technol. 83 (1), 247–250. doi:10.2166/wst.2020.569.

Insel, G., Ozyildiz, G., Okutman-Tas, D., Guven, D., Zengin, G.E., Pala-Ozkok, I., Kurt, E.S., Atli, E., Artan, N., Takács, I., Cokgor, E., 2022. A comprehensive evaluation of process kinetics: a plant-wide approach for nutrient removal and biogas production. Water Res. 118410. doi:10.1016/j.watres.2022.118410.

Jensen, P.D., Hardin, M.T., Clarke, W.P., 2009. Effect of biomass concentration and inoculum source on the rate of anaerobic cellulose solubilization. Bioresour. Technol. 100 (21), 5219–5225. doi:10.1016/j.biortech.2009.05.018.

Koch, K., Hafner, S.D., Weinrich, S., Astals, S., 2019. Identification of critical problems in biochemical methane potential (BMP) tests from methane production curves. Front. Environ. Sci. 7, 178. doi:10.3389/fenvs.2019.00178.

Koch, K., Hafner, S.D., Weinrich, S., Astals, S., Holliger, C., 2020. Power and limitations of biochemical methane potential (BMP) tests. Front. Energy Res. 8. doi:10.3389/fenrg.2020.00063.

Li, Y., Jin, Y., Li, J., Li, H., Yu, Z., Nie, Y., 2017. Effects of thermal pretreatment on degradation kinetics of organics during kitchen waste anaerobic digestion. Energy 118, 377–386. doi:10.1016/j.energy.2016.12.041.

Motulsky, H.J., Ransnas, L.A., 1987. Fitting curves to data using nonlinear regression: a practical and nonmathematical review. FASEB J. 1 (5), 365–374. doi:10.1096/fasebj.1.5.3315805.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models Part I — a discussion of principles. J. Hydrol. 10 (3), 282–290. doi:10.1016/0022-1694(70)90255-6.

Rittmann, B.E., McCarty, P.L., 2001. Environmental Biotechnology: Principles and Applications. McGraw-Hill, Boston.

Yates, G.T., Smotzer, T., 2007. On the lag phase and initial decline of microbial growth curves. J. Theor. Biol. 244 (3), 511–517. doi:10.1016/j.jtbi.2006.08.017.