Smoothing spline assessment of the accuracy of enteric hydrogen and methane production measurements from dairy cattle using various sampling schemes

Henk J. van Lingen,1*† James G. Fadel,1 Ermias Kebreab,1 André Bannink,2 Jan Dijkstra,3 and Sanne van Gastelen2

1Department of Animal Science, University of California–Davis, Davis, CA 95616
2Wageningen Livestock Research, Wageningen University & Research, 6700 AH, Wageningen, the Netherlands
3Animal Nutrition Group, Wageningen University & Research, 6700 AH, Wageningen, the Netherlands

ABSTRACT

Estimating daily enteric hydrogen (H2) and methane (CH4) emitted from dairy cattle using spot sampling techniques requires accurate sampling schemes. These sampling schemes determine the number of daily samplings and their intervals. This simulation study assessed the accuracy of daily H2 and CH4 emissions from dairy cattle using various sampling schemes for gas collection. Gas emission data were available from a crossover experiment with 28 cows fed twice daily at 80% to 95% of the ad libitum intake, and an experiment that used a repeated randomized block design with 16 cows twice daily fed ad libitum. Gases were sampled every 12 to 15 min for 3 consecutive days in climate respiration chambers. Feed was fed in 2 equal portions per day in both experiments. Per individual cow-period combination, generalized additive models were fitted to all diurnal H2 and CH4 emission profiles. Per profile, the models were fitted using the generalized cross-validation, REML, REML while assuming correlated residuals, and REML while assuming heteroscedastic residuals. The areas under the curve (AUC) of these 4 fits were numerically integrated over 24 h to compute the daily production and compared with the mean of all data points, which was considered the reference. Next, the best of the 4 fits was used to evaluate 9 different sampling schemes. This evaluation determined the average predicted values sampled at 0.5, 1, and 2 h intervals starting at 0 h from morning feeding, at 1 and 2 h intervals starting at 0.5 h from morning feeding, at 6 and 8 h intervals starting at 2 h from morning feeding, and at 2 unequally spaced intervals with 2 or 3 samples per day. Sampling every 0.5 h was needed to obtain daily H2 productions not different from the selected AUC for the restricted feeding experiment, whereas less frequent sampling had predictions varying from 47% to 233% of the AUC. For the ad libitum feeding experiment, sampling schemes had H2 productions from 85% to 155% of the corresponding AUC. For the restricted feeding experiment, daily CH4 production needed samplings every 2 h or shorter, or 1 h or shorter, depending on sampling time after feeding, whereas sampling scheme did not affect CH4 production for the twice daily ad libitum feeding experiment. In conclusion, sampling scheme had a major impact on predicted daily H2 production, particularly with restricted feeding, whereas daily CH4 production was less severely affected by sampling scheme.

Key words: restricted feeding, ad libitum feeding, cow, diurnal profile

INTRODUCTION

In vivo measurements of enteric H2 and CH4 emission from cattle have been published abundantly in the recent scientific literature. Methane emission measurements aim at identifying the effectiveness of CH4 mitigation strategies for animal agriculture (e.g., Hammond et al., 2016a; Patra, 2016), whereas H2 production measurements contribute to increased understanding of rumen fermentation dynamics and metabolic physiology of microbes (e.g., Olijhoek et al., 2016; van Lingen et al., 2017, 2021). Enteric H2 and CH4 emissions from cattle are measured using equipment including the GreenFeed system and climate respiration chambers (CRC). The GreenFeed system is a measurement device for sampling gas emissions over short periods (typically 5–7 min several times within a day), whereas sampling takes place repeatedly up to multiple times per hour in CRC. Patra (2016) stated measurement systems are likely more suitable for evaluation of CH4 mitigation measures if observations are obtained at...
different times of the day relative to the diurnal CH4 emission cycle. Similarly, Hammond et al. (2016a) recommended sampling schemes for gas collection should include sufficient sampling times to account for diurnal and postprandial variation in CH4 emission. Lee et al. (2017) used data from indirect calorimetry systems to simulate spot sampling and to evaluate the accuracy of estimates of gaseous exchanges (O2 consumption, CH4 and CO2 production) of 3 experiments (2 ad libitum feeding, 1 restricted feeding). They concluded spot sampling with a frequency of at least 8 time points (to represent sampling every 3 h within a 24-h cycle) is required to provide accurate estimates of gaseous exchange. Given the greater relative postprandial variation in H2 emission compared with CH4 emission (van Lingen et al., 2017), concerns about the impact of sampling schemes may apply even more to H2 emissions measurements.

Compared with the reported variation in CH4 emission from cattle across studies, variation in H2 emissions appears considerably large. Part of this variation is due to dietary supplements such as 3-nitrooxypropanol (3-NOP; Hristov et al., 2015) and nitrate (Olijhoek et al., 2016) that have a strong potential to increase H2 emissions. However, differences remain large across control treatments of studies. For example, reported daily production of H2 for control treatments was 0.004 g/d (emissions from steers measured using CRC, sampling scheme not reported; Martinez-Fernandez et al., 2018), 0.02 g/d (emissions from dairy cows measured at 8 points staggered over a 3-d period using a GreenFeed system; Hristov et al., 2015), 1.54 g/d (emissions from steers measured every 3 h using CRC; Vyas et al., 2016), and 2.34 g/d (emissions from dairy cows measured on average 5.5 times per day using a GreenFeed system; van Gastelen et al., 2022). Despite this large variation up to almost a factor 600 when comparing the lowest to the highest of these H2 production values, DMI ranged from 6.68 to 28.0 kg/d, which is an increase by only about a factor of 4. Having noted the great relative postprandial variation in H2 emission (van Lingen et al., 2017), this mismatch of relative ranges of H2 emissions and DMI warrants a solid accuracy assessment of sampling schemes of H2 emission measurements. The aim of the present simulation study is to model diurnal variation of daily H2 and CH4 emission from dairy cattle and assess the accuracy of various sampling schemes for estimating daily emission rates. Sampling scheme might not be a completely exhaustive explanation of this mismatch, but for reasons of simplicity, factors such as basal diet composition were not considered. We hypothesize that the more pronounced diurnal patterns in CH4 and H2 production with restricted compared with ad libitum feeding require more frequent sampling to obtain accurate estimates of enteric CH4 and H2 production. Furthermore, we hypothesize that various H2 daily emission rates reported in the literature underestimated the true emission rates, which is at least partly due to insufficiently frequent gas samplings.

MATERIALS AND METHODS

Because no human or animal subjects were used, this analysis did not require approval by an Institutional Animal Care and Use Committee or Institutional Review Board.

Overall Simulation Approach

The overall modeling strategy was to evaluate 4 model fits for H2 and 4 model fits for CH4 using all the data from each experiment and compare these models to a reference, which was the mean of all data points for H2 and then CH4. The model fit that most closely matched reference data would then be selected and considered a reference fit. This reference fit would in turn be compared against 9 sampling schemes. The sampling schemes that may be applied across a range of potential experimental conditions were as few as twice a day up to twice per hour. All the H2 and CH4 data came from 2 experiments with 2 dietary treatments.

Data Sources

Hydrogen and CH4 emissions were collected from a 2 × 2 crossover in vivo experiment with 28 dairy cows (experiment A; van Lingen et al., 2017, 2 × 4 observations from cannulated cows; van Gastelen et al., 2017, 2 × 24 observations from noncannulated cows) and a repeated block design experiment with 16 cows and 4 periods per cow (experiment B; van Gastelen et al., 2020) using CRC. Experiment A consisted of cows fed control versus linseed oil supplemented diets, whereas experiment B consisted of cows that were offered either a placebo diet or a diet supplemented with CH4 inhibitor 3-NOP. Cows were fed 80% to 95% of their ad libitum DMI in experiment A, whereas cows were fed ad libitum in experiment B. The feed was offered in 2 equal portions delivered at 0600 and 1600 h in experiment A and at 0530 and 1530 h in experiment B. For experiment A, DMI was 18.0 ± 0.46 and 18.1 ± 0.62 kg/d for control and linseed oil supplemented diets, respectively, in van Lingen et al. (2017), whereas this was 18.0 ± 0.69 and 17.5 ± 0.69 kg/d for control and linseed oil supplemented diets, respectively, in van Gastelen et al. (2017). For experiment B, DMI was 20.1 ± 0.71 and 20.5 ± 0.71 kg/d for the placebo and 3-NOP treatments, respectively (van Gastelen et al., 2020). The dietary crude fat.
contents were 34 and 56 g/kg DM for the control and linseed oil supplemented diets, respectively, in experiment A. The realized inclusion rate of 3-NOP in the treatment diet in experiment B was 51 mg/kg DM. For additional descriptive statistics of these 2 experiments, we refer to the 3 original publications and Supplemental Table S1. (Supplemental documents are available at https://data.mendeley.com/datasets/mrdrrhbs8/1.) In both experiments, gas emissions of H₂ and CH₄ were recorded for 3 full days using CRC. Exhaust air of the 4 chambers was sampled consecutively per chamber throughout a 12-min period. Every fifth sampling cycle was extended to 15 min for sampling the inlet air. Further details about the experimental setup of the CRC were as described by van Gastelen et al. (2015), except that gas measurements recorded during feeding and milking when animal caretakers were present in the CRC (approximately 30 min, twice daily) were not excluded. The peak H₂ emissions rate of the diurnal cycle is commonly reached during the first 30 min after feed delivery, and excluding these measurements of this 30-min period substantially underestimates the daily H₂ emissions rate.

**Data Processing**

All 56 diurnal H₂ and CH₄ emission profiles for experiment A belonging to the previously mentioned 2 × 4 + 2 × 24 observations were used for analysis. Note that in this paper, a diurnal gas emission profile is the instantaneous gas emission rate (mass per unit of time, here in g/h) throughout a 24-h period. This diurnal emission profile is quantified by fitting the gas emission rate to the corresponding time on the 24-h clock. Measurements of instantaneous gas emission rates collected throughout a 72-h period were considered relative to feed delivery time using the 24-h clock. In other words, time points representing emission rates against time from morning feeding for 3 different days were jointly used for fitting one smoothing spline over 24 h. For experiment A, 99 ± 0.8% of the H₂ measurements were positive and retained across all diurnal profiles. For experiment B, data from 5 diurnal emission profiles from 3 cows were discarded because 1 cow was removed from the experiment after an accident and 2 other cows had health problems that seemed unrelated to the dietary treatments or other experimental conditions. In addition, 55 ± 22.5% of the data points of 4 H₂ diurnal emission profiles recorded for the first 4 cows in experiment B were negative. Therefore, these 4 diurnal emission profiles were excluded from the data, along with a fifth diurnal emission profile of another cow for which 37% of the data was negative. Note that the mean ± standard deviation refers to percentages of data points per diurnal emission profile. Based on the 5 incidences of compromised health of cows that did not overlap with the 5 H₂ emission profiles that contained a high incidence of negative values, 54 H₂ and 59 CH₄ diurnal emission profiles were retained for experiment B. Finally, the remaining negative H₂ data points of the retained diurnal profiles were discarded, which resulted in 97 ± 4.4% (minimum 83%) of the data points retained across the 54 diurnal emission profiles. The excluded negative H₂ data points were mostly recorded during the late-postprandial state at which the H₂ emission rate is low. Therefore, the effect of excluding these negative H₂ data points was considered negligible.

**Fitting Generalized Additive Models**

Values of H₂ emission rate (g/h) had large variation and were log-transformed to stabilize variance before the data were analyzed, whereas the raw CH₄ emission rates were used given the more stable variance of this variable. Per diurnal profile of every individual cow, gas emission rates were analyzed using the following generalized additive model shown in Equation 1:

\[
y_{ij} = \beta_0 + f(t_{ij}) + \epsilon_{ij},
\]

where \(y_{ij}\) denotes the observed gas emission rates of H₂ or CH₄ (g/h per animal) at time from morning feeding \(t_{ij}\) continuous over 0 to 24 h, for measurement \(i = \{1, \ldots, n\}\) on day \(j \{1, 2, 3\}\) with \(n\) the total number of observations of the diurnal emission profiles, \(\beta_0\) denotes the intercept, \(f(t_{ij})\) is a smooth function expanded from a cyclic cubic spline basis function with 70 equidistant knots, and \(\epsilon_{ij}\) denotes the residual error. Note that the data from the 3 consecutive days were pulled together into a single 24-h time frame and a single diurnal gas emission profile, which assumes the 3 d are fully intercomparable. This statistical model was subject to the penalized least squares objective function shown in Equation 2:

\[
\sum_{i=1}^{n} \left( y_{ij} - f(t_{ij}) \right)^2 + \lambda \int_{a}^{b} \{ f''(t) \}^2 dt.
\]

This objective function is an extension of the commonly applied least squares problem with a penalty term in which the integrated squared second derivative of the smooth function is multiplied by a smoothness parameter \(\lambda\), with \(a\) and \(b\) representing the 0 and 24 h boundaries, respectively. The smoothness parameter \(\lambda\) controls the wiggliness penalty and implies a trade-off between smoothness and model fit (e.g., Simpson, 2018; Perperoglou et al., 2019). Herewith, the value of \(\lambda\)
rather than the knot selection controls the smoothness of the fit, with \( \lambda = 0 \) indicating no penalty and \( \lambda = \infty \) indicating a linear fit. A generalized additive model was fitted for which \( \lambda \) was estimated using a generalized cross-validation (GCV). After fitting this model to all \( \text{H}_2 \) and \( \text{CH}_4 \) diurnal emission profiles, outliers were identified using model residuals. The outlier identification was performed using the interquartile range method (e.g., Zwilinger and Kokoska, 2000) with a factor of 3.0 for extremes in constructing boundaries, after which records of residuals outside these boundaries were excluded from the data. Throughout the diurnal emission profiles, 100% of the \( \text{H}_2 \) and \( \text{CH}_4 \) measurements in experiments A and B were retained.

After the removal of outliers from the data, model fitting was performed again by estimating \( \lambda \) (i.e., the smoothness selection) using GCV, or treating the smooth as a random effect in which \( \lambda \) is a variance parameter that was estimated using the REML. Four alternative model fits, namely 1 GCV and 3 REML fits, were considered for both the \( \text{H}_2 \) and \( \text{CH}_4 \) diurnal emission profiles: 1) the residuals were considered independent according to \( \varepsilon_{ij} \sim N(0, \sigma^2) \) and \( \lambda \) was estimated by GCV; 2) the residuals were considered independent according to \( \varepsilon_{ij} \sim N(0, \sigma^2) \) and \( \lambda \) was estimated by REML; 3) the residuals were considered correlated due to repeated measurements according to \( \varepsilon_{ij} \sim N(0, \Lambda \sigma^2) \), with \( \Lambda \) a rational quadratic spatial correlation matrix, and \( \lambda \) was estimated by REML (REMLc fit); 4) the residuals were considered heteroscedastic according to \( \varepsilon_{ij} \sim N(0, \exp(2\delta y_{ij}) \sigma^2) \), where \( \delta \) is a variance parameter that needs to be estimated, and \( \lambda \) was estimated by REML (REMLv fit). After analyzing diurnal emission profiles of individual cows, diurnal emission profiles were analyzed per experiment and dietary treatment. The latter approach, which should be regarded as a summary of the fitting of the individual diurnal emission profiles, indicated the average gas emission rates of all cows per dietary treatment and experiment were estimated. The generalized additive model shown in Equation 1 was applied again using GCV for estimating \( \lambda \), but in this case \( i = \{1, \ldots, n\} \) indicated the total number of observations of all diurnal emission profiles per dietary treatment and experiment. All generalized additive models were fitted using the mgcv package (Wood, 2017) in R version 3.6.3 (R Core Team, 2019).

### Evaluating Daily Gas Emission

After fitting the generalized additive models (i.e., the smoothing splines), areas under the curve (AUC) for every of the 4 model fits per observed diurnal gas emission profile were calculated by numerical integration of each fit over 24 h using a step size of 0.01 h. Numerical integration was performed by taking predicted values of each model fit that were obtained using the predict(.) function in R, after which the means of the predicted values were multiplied by 24 h. Per diurnal emission profile, the mean of the observed gas emission rates (g/h) multiplied by 24 h was taken as reference production (g/d) to evaluate the accuracy of AUC of the 4 different model fits. Subsequently, predicted values of \( \text{H}_2 \) or \( \text{CH}_4 \) emission were sampled from the model fit for which the AUC was the closest to the reference. The following 9 sampling schemes applied to a 24-h period were then considered for simulation: a) every 8 h starting at 2 h after morning feeding, b) every 6 h starting at 2 h after morning feeding, c) every 2 h starting at morning feeding, d) every 2 h starting at 0.5 h after morning feeding, e) every hour starting at 0.5 h after morning feeding, f) every hour starting at morning feeding, and g) every 0.5 h starting at morning feeding. Sampling schemes reflecting a draw every x.x h starting at y.y h from morning feeding time were written as x.x_y.y (e.g., 8.0_2.0 for sampling scheme a). Given the peak \( \text{H}_2 \) emission rate around 0.5 h after feed delivery (van Lingen et al., 2017), these sampling schemes have starting points before, at or after 0.5 h, and then either include or exclude the \( \text{H}_2 \) peak emission rate. In addition to these schemes with equally spaced samplings, 2 other sampling schemes that represented unequally spaced sampling were considered for simulation (based on Hammond et al., 2015), h) sampling at 1.0, 7.0, and 16.0 h after morning feeding (UnEq3), and i) sampling at 1.0 and 7.0 h after morning feeding (UnEq2). Every sampling scheme enabled the calculation of daily \( \text{H}_2 \) and \( \text{CH}_4 \) production (g/d) for every diurnal emission profile by taking the mean of the sampled values multiplied by 24 h. The AUC, the calculated daily emissions using the various sampling schemes along with the reference were evaluated using the statistical model in Equation 3:

\[
y_{ijkmn} = x_{ijkmn}^T \beta + a_i + \varepsilon_{ijkmn},
\]

where \( y_{ijkmn} \) is the \( i \)th daily gas emission rate (g/d) that was calculated using the \( j \)th way of sampling (\( j \) represents either the reference and the 4 AUC, or 1 selected AUC as the reference model and 9 sampling schemes), the \( k \)th diet, \( l \)th period, \( n \)th respiration chamber, and for the \( m \)th cow; \( x_{ijkmn}^T \) is the transpose vector of explanatory variables representing sampling, diet, the 4 different respiration chambers, and the experimental period; \( \beta \) is a vector containing the coefficients that quantify the effect sizes associated with the explanatory
variables in $x_{ijhmn}^{T}$, including the interaction between sampling scheme and dietary treatment; $a_i$ is a vector of random cow effects; and $\varepsilon_{ijhmn}$ represents the residuals. Models assumed homoscedastic residual variance according to $\varepsilon_{ijhmn} \sim \left(0, \sigma_e^2\right)$ or heteroscedastic residual variance according to $\varepsilon_{ijhmn} \sim \left(0, \delta_p^2 \sigma_e^2\right)$, with $\delta = 1$ and $\delta_p$ to be estimated; $\sigma_e^2$ is the residual variance. Stratification variables for $\delta_p$ were sampling, diet, or the sampling × diet interaction, with $p$ the number of samplings, diets, or samplings × diets. Final model selection was based on the Akaike information criterion (AIC; note that AIC was not used for the smoothness selection of the generalized additive models). Normality and homoscedasticity of the residuals were visually inspected with residuals versus predicted plots. The variance homoscedasticity assumption was violated for the H$_2$ emissions of experiment B, despite fitting a model that did not assume homoscedasticity, but could be solved by removing the random effect of cow from the model. After fitting the models, daily H$_2$ and CH$_4$ emission rates for the 9 sampling schemes were evaluated using multiple comparisons according to the Tukey-Kramer method. Significant differences were declared at $P \leq 0.05$, and trends toward significance at $0.05 < P \leq 0.10$. Model fitting, multiple comparisons, and calculation of the estimated marginal means were performed using the nlme (Pinheiro et al., 2020), multcomp (Hothorn et al., 2008), and emmeans (Lenth, 2020) packages, respectively, in R version 3.6.3.

RESULTS

Hydrogen Emission

Estimating the H$_2$ emission rates per dietary treatment and experiment fitted by models using GCV indicated a relatively sharp postprandial peak of emissions for experiment A in which cows were fed restrictedly. However, a more moderate peak was observed for both dietary treatments of experiment B in which cows were fed ad libitum, despite the higher emission rate for the 3-NOP than the placebo treatment (Figure 1). Estimating the individual H$_2$ emission rates of experiment A with sharp postprandial peaks by REML and GCV resulted in relatively wiggly fits, whereas smoother fits were obtained for REMLe and occasionally for REMLv (Supplemental Figure S1, https://data.mendeley.com/datasets/2rhhbas81/). These fits resulted in AUC of which the wigglier GCV fits underestimated the daily emission rate quantified by the reference the least (1.45 ± 0.08 vs. 1.67 ± 0.08 g/d; Figure 2a), and the other 3 AUC resulted in larger (REMLc) or numerically larger underestimations (REML, REMLv) compared with GCV. Therefore, the GCV model was selected as the reference model to compare against the 9 sampling schemes. No interactions between sampling and dietary treatment were observed, hence the results below apply regardless of dietary treatment. The 0.5_0.0 sampling scheme, which sampled from the GCV fits every 0.5 h starting at 0 h, was the only sampling scheme that did not result in a difference from the AUC of the GCV fits (Figure 3a). Applying the 8.0_2.0, 6.0_2.0, 2.0_0.0, and 1.0_0.0 sampling schemes resulted in underprediction with an absolute emission rate down to 0.68 ± 0.05 g/d, whereas the application of 1.0_0.5, 2.0_0.5, UnEq2, and UnEq3 resulted in overprediction with an absolute daily H$_2$ emission up to 3.36 ± 0.16 g/d. These extremes then ranged from 47% to 233% of the AUC GCV fits.

Similarly as for experiment A, modeling the H$_2$ emission rates of experiment B with more moderate peaks (placebo and 3-NOP) by REML and GCV resulted in relatively wiggly fits, whereas REMLe fits were commonly smoother (Supplemental Figure S2, https://data.mendeley.com/datasets/2rhhbas81/). Fits obtained by REMLv did not appear as smooth as the REMLe fits, but had higher gas emission rates in various cases and also deviated from the REML and GCV fits in several cases. The AUC of diurnal gas emission profiles fitted by REMLv were the closest to the reference, and this fit was selected to sample from using the 9 sampling schemes. Comparing the reference, the REMLv AUC and the emission rates sampled from the REMLv AUC resulted in a significant interaction between sampling and dietary treatment (placebo and 3-NOP). When evaluating observations of the placebo treatments only, the REMLv fits were not significantly different from the reference (0.71 ± 0.20 vs. 0.78 ± 0.19 g/d; Figure 2b), whereas the other 3 AUC underestimated the reference, of which REMLe fits had the lowest H$_2$ emission rate (0.52 ± 0.19 g/d). Sampling from the REMLv fits indicated the UnEq2 and UnEq3 sampling scheme resulted in higher daily H$_2$ emission (1.14 ± 0.09 and 0.99 ± 0.06 g/d, respectively; Figure 3b) than the REMLv AUC, whereas the 6.0_2.0 sampling scheme had a lower H$_2$ emission rate (0.62 ± 0.02 g/d). These extremes then ranged from 85% to 155% of the AUC REMLv fits. Furthermore, the 2.0_0.0 and 2.0_0.5 sampling schemes had emission rates of 0.65 ± 0.02 g/d versus 0.86 ± 0.05 g/d, respectively, which were different from each other. When evaluating observations of the 3-NOP treatments only, the AUC of the REMLv fits showed the greatest underestimation of the reference again (4.65 ± 0.26 vs. 8.04 ± 0.27 g/d), whereas REMLv still had a daily H$_2$ emission rate that was not different from the reference (7.37 ± 0.83 g/d;
Figure 1. Smoothing spline fits of H₂ and CH₄ emissions (g/h) per experiment and dietary treatment by a generally additive model fitted using generalized cross-validation. Gray dots represent individual observations; vertical dashed lines indicate times of feed delivery. 3-NOP = 3-nitrooxypropanol.
Figure 2c). Then, sampling from these REMLv fits indicated no significantly different daily H₂ emissions across the sampling schemes (Figure 3c), although the UnEq2 and UnEq3 sampling schemes had H₂ emission rates of 11.0 ± 1.29 and 10.4 ± 1.11 g/d, respectively, which numerically overestimated the REMLv by more than 3 g/d.

**Methane Emission**

Estimating the CH₄ emission rates per dietary treatment and experiment indicated a relatively sharp post-prandial peak of emissions for experiment A in which cows were fed restrictedly, whereas a more moderate peak was observed for both dietary treatments of experiment B, despite the lower emission rate for the 3-NOP than the placebo treatment (Figure 1). The CH₄ emission rate for the 3-NOP treatment first declined after feeding, after which the emission rate reached its peak value. This decline before the peak suggests 3-NOP is effective before the CH₄ production rate is upregulated due to carbohydrate fermentation and subsequent methanogenesis. The GCV, REML, REMLv, and REMLc fits for the individual CH₄ emission rates of experiment A did not show major differences, although the GCV fits showed slightly wigglier behavior in some cases (Supplemental Figure S3, https://data.mendeley.com/datasets/mrdrrhbf8/1). Despite these relatively similar fits, the daily CH₄ emission rates calculated from the AUC for the GCV and REMLv fits (381.6 ± 9.5 and 370.2 ± 9.5 g/d, respectively; Figure 4a) were different from the reference CH₄ emission rate (388.2 ± 9.5 g/d). When using 3 decimals for the estimated marginal means, the REMLc AUC were numerically the closest to the reference and these fits were used for evaluating the sampling schemes. No interactions between sampling and dietary treatment were observed and, therefore, there was no need to evaluate the daily CH₄ emission rates for the 2 treatments separately. Daily CH₄ emission rates for the 0.5_0.0, 1.0_0.0, 1.0_0.5, and 2.0_0.0 sampling schemes were not significantly different from each other and from the AUC of the REMLc fits (Figure 5a). However, the 2.0_0.5, 6.0_2.0, 8.0_2.0, UnEq2, and UnEq3 sampling schemes all significantly overestimated the REMLc AUC, of which the UnEq2 sampling scheme had the greatest CH₄ emission rate of 425.3 ± 9.6 g/d, which is 111% of the AUC REMLc fit.

Although the CH₄ emission rates in experiment B for cows fed ad libitum showed relatively flat diurnal patterns compared with experiment A, the GCV fits for experiment B occasionally showed relatively wiggly fits compared with the REMLv fits (Supplemental Figure...
Figure 3. Hydrogen emission (estimated marginal mean ± SE; g/d) from dairy cows. Emissions were calculated from the area under the curve and by sampling every x.x h starting at y.y h from morning feeding time (x.x_y.y), at 1.0, 7.0, and 16.0 h after morning feeding (UnEq3), and at 1.0 and 7.0 h after morning feeding (UnEq2) from the selected fit. Hydrogen emissions are shown for (a) all diurnal profiles of experiment A, in which cows were fed at 80% to 95% of ad libitum intake and generalized cross-validation (GCV) fits were used for sampling; (b) the diurnal profiles for all cows in experiment B that received the placebo diet and REML fitted smoothing spline models that assumed heteroscedastic residual variance (REMLv) fits were used for sampling; and (c) the diurnal profiles for all cows in experiment B that received the 3-nitrooxypropanol diet and REMLv fits were used for sampling. Values with a different letter indicate a significant difference (P < 0.05).
The AUC of this relatively smooth REMLv fit resulted in daily CH$_4$ emission rates that were significantly lower than the reference (366.0 ± 12.1 g/d; Figure 4b), whereas the other 3 AUC were not different from the reference CH$_4$ emission rate. The REML AUC were numerically the closest to the reference (384.1 ± 12.1 g/d). Therefore, the REML model was selected as the reference model to compare against the 9 sampling schemes. No interactions between sampling and dietary treatment were observed. None of the sampling schemes had daily CH$_4$ production rates different from the REML AUC (Figure 5b). The UnEq2 sampling scheme had an estimated marginal mean CH$_4$ emission rate (376.5 ± 12.1 g/d; Figure 5b), which was not significantly less than obtained for the REML AUC, but was significantly less than obtained for the UnEq3 and 6.0_2.0 sampling schemes. All other sampling schemes had daily CH$_4$ production rates that were not significantly different from each other.

**DISCUSSION**

To the best of our knowledge, the present simulation study is the first to quantify the implications of various sampling frequencies and sampling times on the estimated daily H$_2$ rates of dairy cattle, and the first on the estimated CH$_4$ emission rates using smoothing splines. The results of the present paper illustrate that sampling schemes can have a large impact on the observed daily H$_2$ emission rates, whereas daily CH$_4$ emission rates seem affected by sampling scheme to a lesser extent. The present study is particularly relevant for studies using spot sampling techniques that aim at measuring H$_2$ emissions to increase our understanding of the dynamics of rumen fermentation and at measuring CH$_4$ emissions to evaluate the effectiveness of strategies to mitigate greenhouse gas emissions from cattle. Conclusions drawn from the observed results regarding required sampling frequencies and sampling times relative to a reference model may apply to GreenFeed systems, in particular, but may also be of use for measurement techniques such as CRC. Our database contained 2 studies with different diurnal gas emission profiles that could be considered as extremes given the different feeding regimens and the dietary supplementation with 3-NOP, which potently increases H$_2$ and decreases CH$_4$ emission rates.

**Generalized Additive Model Approach**

The application of smoothing splines is intrinsically associated with large flexibility, which means they follow the true emission rate closely and give accurate expected gas emissions rates. Replacing a smoothing spline by a certain nonlinear forcing function would take this flexibility away and then likely affect the estimated daily gas emission rates. Although experimental conditions are often such that the feeding regimen is

---

**Figure 4.** Methane (CH$_4$) emission (estimated marginal mean ± SE; g/d) from dairy cows that was calculated by taking the mean of the measured values in climate respiration chambers (Ref) and by calculating the areas under the curve of smoothing spline models fitted using generalized cross-validation (GCV) and REML. REML fitted smoothing spline models that assumed correlation of repeated measurements (REMLc), or heteroscedastic residual variance (REMLv). The CH$_4$ emissions are shown for (a) all diurnal profiles of experiment A, in which cows were fed at 80% to 95% of ad libitum intake, and (b) the diurnal profiles for all cows in experiment B that received the placebo diet and were fed ad libitum. Values with a different letter indicate a significant difference ($P < 0.05$).
relatively tight, the intake of feed by the animal may not be tightly controlled, which could result in one or more gas emission peaks between 2 moments of feed delivery. A smoothing spline would then follow the true emission rate more closely, whereas a nonlinear forcing function assumes a certain fixed pattern of gas emission rate that could deviate from the true emission rate.

**Predicted Values of H₂ and CH₄ Emission Rates**

In the present study, predicted values of fitted smoothing splines were sampled for calculating daily H₂ and CH₄ production (g/d). Selecting a fitted smoothing spline per diurnal emission profile, and subsequently sampling predicted values, results in variant gas emis-

---

**Figure 5.** Methane (CH₄) emission (estimated marginal mean ± SE; g/d) from dairy cows. Emissions were calculated from the area under the curve and by sampling every x.x h starting at y.y h from morning feeding time (x.x_y.y), at 1.0, 7.0, and 16.0 h after morning feeding (UnEq3), and at 1.0 and 7.0 h after morning feeding (UnEq2) from the selected fit. The CH₄ emissions are shown for (a) all diurnal profiles of experiment A, in which cows were fed at 80% to 95% of ad libitum intake, REML fitted smoothing spline models that assumed correlation of repeated measurements (REMLc) fits were used for sampling; and (b) the diurnal profiles for all cows in experiment B that received the placebo diet and generalized cross-validation (GCV) fits were used for sampling. Values with a different letter indicate a significant difference ($P < 0.05$).
sion rates due to the different diurnal profiles from different cows, whereas these sampled values do not contain errors due to predicted values being error free (e.g., St-Pierre, 2003). Taking predicted values, therefore, leaves out any residual variance around the true emission rates of H₂ and CH₄, which may have reduced the variance of the daily H₂ and CH₄ production that was calculated. However, it is still questionable if this omission of residual variance had a substantial effect on daily H₂ and CH₄ production and its variance. The expected residual variance will always be zero, which means that the expected values of the calculated daily H₂ and CH₄ production and the estimated marginal means will not be affected. If the variance of the daily H₂ and CH₄ production is seriously affected, the significance of the multiple comparisons will also be affected. The true variances of the daily H₂ and CH₄ production could be determined by taking the mean and variance of the raw data points measured close to the predetermined sampling schemes. However, an advantage of the present analysis is that all data are used rather than a few points matching the sampling schemes. In another approach, Lee et al. (2022) used observed CH₄ emissions measured at selected time points according to various sampling frequencies (every 2, 3, 4, or 6 h in a 24-h cycle). Results in that study, therefore, suffer from the drawbacks described before. Nonetheless, Lee et al. (2022) recommended sampling at least every 3 h per 24-h cycle to estimate daily CH₄ production.

The selection of the GCV, REML, REMLv, or REMLc fit by comparing to a reference obtained by taking the mean of all data points may not be a commonly used model selection method. The AIC and Bayesian information criterion (BIC) are the commonly accepted fit statistics for model selection (e.g., Buscemi and Plaia, 2020), but were not used for selecting the smoothness of the generalized additive models in the present study. Model selection using AIC and BIC could possibly affect the gas emission rates obtained using the various sampling schemes as well. However, when comparing the GCV and REML fits for H₂ and CH₄ emission rates of experiments A and B, AIC consistently favored the GCV fits, whereas the BIC consistently favored the REML fit. These contradicting information criteria then left the model selection undecided, which was then solved by comparing the different fits to a reference. Furthermore, AIC and BIC favored various fits that oversmoothed diurnal emission profiles. These oversmoothings were identified by the residual plots of favored fits that showed certain patterns. For example, these favored oversmoothings were found for REMLv versus REML fits for H₂ and CH₄ of both experiments and for REMLc versus REML fits for H₂ in both experiments. Therefore, AIC and BIC were not used for selecting generalized additive model fits in the present study. Finally, it should be noted that the chosen approach and model selection may have implications for gas emission rates obtained for the sampling schemes. For example, applying the sampling schemes to GCV fits rather than REMLv fits of the H₂ diurnal emission profiles of experiment B affected the daily H₂ emission rates that were obtained (result not shown). The 2.0_0.0 and 6.0_2.0 sampling schemes both significantly underestimated the H₂ emission rate for the placebo treatments. These 2 sampling schemes had significantly less H₂ emission for the 3-NOP treatments as well. Therefore, our results on the impact of sampling scheme on emissions depend partly on the choice of reference model.

**Methane Emissions**

Enteric CH₄ emission measurements may be inaccurate for various reasons (Hammond et al., 2016a), but based on our results, sampling schemes do not significantly contribute to inaccurate measurements for the experiment with twice daily ad libitum feeding, given that at least 3 measurements per day are taken. It should be noted, though, that the method used for evaluating measurement accuracy may have an effect. Lee et al. (2022; 2 experiments with ad libitum feeding and one experiment with restricted feeding) concluded at least 8 measurements per day were required using a regression analysis, whereas their multiple comparisons indicated 4 measurements per day would suffice. However, sampling scheme appeared crucial for experiment A in which restricted feeding was applied, likely due to the larger variation in CH₄ emission rate throughout a day. Crompton et al. (2011) observed that the range in hourly CH₄ emission rate during the day decreased with increasing number of feed deliveries per day. The present study indicates twice daily ad libitum feeding did not result in a range in CH₄ emission rates for which more than 3 measurements per day would be required based on sampling from a smoothing spline. However, for restricted feeding, measuring every hour may be necessary for obtaining accurate CH₄ emission rate. When using the GreenFeed system, measurements are often taken at fixed time points for several consecutive days. For example, Hristov et al. (2015) and Roque et al. (2019) performed 8 samplings with on average a 6-h interval throughout a 3-d period. Effectively, their sampling scheme is similar to the 6.0_2.0 sampling scheme that was applied in the present study, which resulted in accurate estimates of the daily CH₄ emissions upon twice daily ad libitum feeding. Therefore, studies such as Hristov et al. (2015) and other studies, which applied twice daily ad libitum feeding and sampled enteric CH₄...
with a similar frequency, have likely obtained accurate estimates of the daily emission rate. However, some studies (e.g., Laubach et al., 2013; Hammond et al., 2016b) report that the CH$_4$ emission rate is the lowest before the first meal of the day, which may point to the need of balancing measurements over day and night. The latter need could be related to feeding and ruminating that may occur during the daylight and overnight hours, respectively (Schirmann et al., 2012).

**Hydrogen Emissions**

The sampling scheme affected the daily H$_2$ emission from experiment A in which restricted feeding was applied rather severely, and affected the daily H$_2$ emission for the placebo treatment from experiment B in which ad libitum feeding was applied to a milder extent. Due to relatively large error estimates, daily H$_2$ emissions for the 3-NOP treatment from experiment B were not affected by sampling, despite the fact that the UnEq3 and UnEq2 sampling schemes had daily emissions that were about 40% greater than observed for the other sampling schemes. A sharper postprandial peak in H$_2$ emissions with restricted feed intake may have resulted from 2 portions of feed intake upon feed delivery (van Lingen et al., 2017), whereas ad libitum feeding and constant feed allowance might have resulted in various portions of feed intake throughout the day. Crompton et al. (2011) observed a decreased range of hourly CH$_4$ emission rate in response to increased number of feed deliveries per day, which qualitatively aligns with the H$_2$ diurnal emission profile for experiments A and B. Therefore, the sampling scheme is crucial for measuring H$_2$ emission rates from diets and feed intake patterns that stimulate sharp peaks in the emission rate in response to feeding. Ad libitum feeding rather than restricted feeding and diets with rather than without a methanogenic inhibitor flattened the peaks of H$_2$ emission profiles. Hourly sampling still not resulting in accurate estimation of the daily gas emission rate (i.e., 1.0_0.5 and 1.0_0.0 sampling for experiment A) indicates that H$_2$ must be sampled every 0.5 h for an accurate estimate of the daily emission rate when feeding restrictedly. Moreover, although feed delivery and availability that affect feed intake patterns can be perfectly controlled, the feed intake rate by the animal as such cannot be controlled. If an animal does not ingest its feed at a later time than at delivery, a delayed and sharp peak in H$_2$ emission is still expected, which points to the need of frequent sampling throughout the day. For ad libitum feeding, hourly sampling appeared to be the minimum frequency required for accurately estimating the daily emission rate for a diet without CH$_4$ inhibitor. Therefore, the use of equipment such as GreenFeed systems and other experimental setups for which this high frequency of sampling is not practically feasible is not suitable for quantifying the daily H$_2$ emission rate of dairy cattle. However, this may not rule out that trends in H$_2$ emission rate in response to different dietary treatments are still accurate (e.g., Vyas et al., 2018; Roque et al., 2019).

The data from the 2 experiments used for the present analysis revealed relatively large differences in H$_2$ emission, with the reference rates being 1.67 and 0.78 g/d for experiment A and the control diet of experiment B, in which restricted and ad libitum feeding was applied that resulted in H$_2$ yields of 0.093 and 0.040 g/kg DMI, respectively. Although the power of 2 individual studies may be limited, this more than doubling in H$_2$ emission rate might be caused by the feeding regimen. In contrast, Veneman et al. (2015) observed H$_2$ yields of 0.038 and 0.060 g/kg DMI for the control treatments of 2 experiments in which restricted and ad libitum feeding was applied, respectively. The latter values may then also suggest that the feeding regimen controls the H$_2$ yield, but in the opposite direction compared with experiments A and B. However, Veneman et al. (2015) excluded the gas measurements recorded during the time the chambers were open twice daily for 45 min for feeding and milking, which is a procedure that was not applied for the present experiments A and B. As shown in the present analysis, restricted feeding requires sampling to take place every 30 min and peaks could occur relatively rapidly after feed delivery. Therefore, the daily H$_2$ emission rate could be underestimated when excluding observations made during milking and feeding. This underestimation has been demonstrated by van Gastelen et al. (2017; experiment A), where excluding the gas concentration measurements during feeding and milking underestimated daily H$_2$ production by 15.2 ± 6.89%, whereas daily productions of CH$_4$ were unaffected. Borsting et al. (2020) also excluded gas measurements recorded during the time the chambers were open. They applied ad libitum feeding, which may have made an underestimation of the H$_2$ emission rate due to less frequent sampling being less likely. For an accurate quantification of the daily H$_2$ emission rate, it is thus not recommended to exclude H$_2$ gas measurements recorded during milking and feeding or when CRC are open. If measurements recorded during these periods cannot be trusted, we recommend the feed to be delivered at the end of the period that the CRC are open, immediately before closing the doors.

Although the 8.0_2.0, 6.0_2.0, and 2.0_0.0 sampling schemes had the lowest daily H$_2$ emission rate in experiment A and the placebo treatment of experiment B, the absolute estimate was never less than 0.62 g/d as obtained for the 6.0_2.0 sampling scheme in experi-
ment B. Therefore, apart from overestimation of daily H\(_2\) production for the UnEq2 and UnEq3 samplings in experiment A, insufficiently frequent sampling could also result in significant underestimation. However, various published studies report emission rates up to about a factor of 100 lower, despite DMI values that were not less than a third of the amount observed for the present experiments A and B and sampling schemes with at least 3 measurements per day. For example, reported H\(_2\) emission rates were as low as 0.02 g/d (Hristov et al., 2015), 0.005 g/d (Lopes et al., 2016), and 0.004 g/d (Martinez-Fernandez et al., 2018). Of these 3 studies, Hristov et al. (2015) and Lopes et al. (2016) used GreenFeed systems, whereas Martinez-Fernandez et al. (2018) used CRC. Nonetheless, other studies that used CRC reported H\(_2\) emission rates of 0.7 to 1.0 g/d (Veneman et al., 2015), 1.0 g/d (Olijhoek et al., 2016), 1.54 g/d (Vyas et al., 2016), and 2.34 g/d (van Gastelen et al., 2022) for control diets, and are in line with observed emission rates for experiment A and the placebo treatment of experiment B. Therefore, it needs to be scrutinized which other factors of the measurement setups affect the detected H\(_2\) concentration. For accurately measuring CH\(_4\), Hammond et al. (2016a) pointed to the accuracy of the collection of air emitted from an animal. Although reasonable values of emitted CH\(_4\) might suggest the collection of emitted air was performed properly, H\(_2\) is substantially lighter than air and may ascend (Atkins and De Paula, 2013) like a balloon filled with helium. Controlled mixing of air flows of the closed environment used for measuring emitted gases from cattle, therefore, may be crucial for the H\(_2\) concentration that is detected. Improper mixing or gas leakages may cause substantial underestimation of the true H\(_2\) emission rate. Therefore, when quantifying H\(_2\) emissions from cattle, it is recommended to verify the quality of the closed environment experimental setup using a H\(_2\) gas recovery test (Gerrits et al., 2018).

**Recommendations for Measuring Enteric Gas Emissions**

The present study has evaluated the implications of applying various spot sampling schemes of enteric gaseous emissions, referring to the number of times per day and the moments of sampling. From the present analysis, one may conclude that sampling 3 times per day at equally spaced time points (e.g., the 8.0_2.0 sampling scheme) suffices for measuring CH\(_4\) emission in case twice daily ad libitum feeding is applied. Given the 3-d measurement period applied for generating the data, this would result in a total of 9 measurements, which is less than 20 measurements taken through-out at least 7 d as recommended by a previous study that used a GreenFeed system (Manafiazar et al., 2016). However, the latter numbers of measurements and measurement days also suggests a threshold of approximately 3 samplings per day, which is in line with Lee et al. (2022). The minimum number of days throughout that need to be sampled may depend on the measurement technique as Arbre et al. (2016) reported a 3-d measurement period sufficed for the SF\(_6\) tracer technique, whereas at least 7 d were needed for a GreenFeed system. It should be noted, though, the present 2 data sets obtained from CRC may not closely resemble all possible experimental circumstances worldwide. Specifically, number of sampling days and animals and measurement techniques were not evaluated in the present study. Nonetheless, the present analysis provides a recommendation regarding the minimum number of samplings per day and their corresponding time points based on 2 experiments differing in feeding management. Sampling schemes with fewer samplings than those that were different from the reference will result in inaccurate measurements, whereas those sampling schemes that were not different from the reference were not proven to be inadequate. Therefore, when measuring CH\(_4\) emissions from cows fed ad libitum at least twice daily, we recommend that researchers sample at least 3 times per day. Threshold frequencies for restricted feeding and measuring H\(_2\) emissions can be extracted similarly from the reported results.

**CONCLUSIONS**

A sampling interval as short as 0.5 h was required for observing a daily H\(_2\) emission rate for experiment A in which restricted feeding and a twice daily feeding frequency was applied. A sampling interval of no more than 2 h was required for estimating the daily H\(_2\) emission rate for the placebo dietary treatment in experiment B that used ad libitum feeding. Sampling scheme did not affect daily H\(_2\) emission for the 3-NOP treatment. At a feeding frequency of twice per day, daily CH\(_4\) production required sampling intervals of 2 h or shorter intervals for the experiment in which restricted feeding was applied, whereas sampling scheme did not affect CH\(_4\) production for the experiment in which twice daily ad libitum feeding was applied. Therefore, sampling scheme had a major impact on predicted daily H\(_2\) production, particularly with restricted feeding, whereas for prediction of daily CH\(_4\) production a more frequent sampling is required with restricted feeding as compared with ad libitum feeding. Other factors contributing to inaccurate quantification of H\(_2\) emission rates remain to be identified.
ACKNOWLEDGMENTS

We acknowledge Maik Kindermann and DSM Nutritional Products (Basel, Switzerland) for permitting the data by van Gastelen et al. (2020) to be transferred from Wageningen Livestock Research to the University of California, Davis. AB and SG acknowledge the support from the a FACCE ERAGAS-funded project CEDERS (co-funded from the Dutch Research Council (NWO project ALW.GAS.2) and the Dutch Public Private Partnership theme Agri & Food). HL designed the research, performed all simulations, and wrote the paper. HL, AB, and JD did the conceptualization. JF, AB, JD, and SG supervised the work. EK was responsible for the project administration. All authors reviewed manuscript drafts and approved the final version. The authors have not stated any conflicts of interest.

REFERENCES


