

Prediction of enteric methane emissions from cattle

LUIS E. MORAES¹, ANDERS B. STRATHE², JAMES G. FADEL¹, DAVID P. CASPER³ and ERMIA S. KEBREAB¹

¹Department of Animal Science, University of California, Davis, CA 95616, USA, ²Department of Basic Animal and Veterinary Sciences, University of Copenhagen, Copenhagen 1870, Denmark, ³Department of Dairy Science, South Dakota State University, Brookings, SD 57007, USA

Abstract

Agriculture has a key role in food production worldwide and it is a major component of the gross domestic product of several countries. Livestock production is essential for the generation of high quality protein foods and the delivery of foods in regions where animal products are the main food source. Environmental impacts of livestock production have been examined for decades, but recently emission of methane from enteric fermentation has been targeted as a substantial greenhouse gas source. The quantification of methane emissions from livestock on a global scale relies on prediction models because measurements require specialized equipment and may be expensive. The predictive ability of current methane emission models remains poor. Moreover, the availability of information on livestock production systems has increased substantially over the years enabling the development of more detailed methane prediction models. In this study, we have developed and evaluated prediction models based on a large database of enteric methane emissions from North American dairy and beef cattle. Most probable models of various complexity levels were identified using a Bayesian model selection procedure and were fitted under a hierarchical setting. Energy intake, dietary fiber and lipid proportions, animal body weight and milk fat proportion were identified as key explanatory variables for predicting emissions. Models here developed substantially outperformed models currently used in national greenhouse gas inventories. Additionally, estimates of repeatability of methane emissions were lower than the ones from the literature and multicollinearity diagnostics suggested that prediction models are stable. In this context, we propose various enteric methane prediction models which require different levels of information availability and can be readily implemented in national greenhouse gas inventories of different complexity levels. The utilization of such models may reduce errors associated with prediction of methane and allow a better examination and representation of policies regulating emissions from cattle.

Keywords: bayesian modeling, climate change, livestock, methane, prediction

Received 13 August 2013; revised version received 27 September 2013 and accepted 4 November 2013

Introduction

The last decade's financial and food price crises, marked by high volatility in food prices, have brought the world population to levels of hunger and undernourishment never reached before (FAO, 2011). The number of undernourished and hungry people reached 1 billion in 2009 and projections for 2010 suggested that at least 925 million people were in similar condition (FAO, 2011). United Nations' forecast indicates that the world population will reach 8.9 billion inhabitants in 2050 (UN, 2004). In this context, an increase in food production, driven by an increase in agriculture productivity, is required to feed the global population. However, the increase in food production may be hampered by society exerting pressure on the livestock production systems due to greenhouse gas emissions, manure

excretion of phosphorus and nitrogen, and water use. Several environmental policies were implemented worldwide in the last two decades, mainly with the objective of reducing mineral and nitrogen excreted by livestock (Oenema, 2004). Most of the attention recently has been directed to emissions of greenhouse gases, especially from ruminants, because methane emissions from enteric fermentation are responsible for 21% of anthropogenic methane emitted in the United States (EPA, 2011).

In order to feed the world population while minimizing environmental impacts of livestock production, a balance must be achieved between an increase in food production with reduced environmental impacts through the identification and examination of more sustainable animal production scenarios. These rely on the prediction of methane emissions at the individual animal level because measurements of daily emissions from livestock require specialized equipment and may be expensive (Johnson & Johnson, 1995). Moreover,

Correspondence: Ermias Kebreab, tel. +530 752 5907, fax: +530 752 0175; e-mail: ekebreab@ucdavis.edu

prediction models of methane emissions are needed for the examination of emission mitigation strategies, environmental policies, the development of decision support models (Moraes *et al.*, 2012), and national greenhouse gas inventories (IPCC, 2006). National inventories of greenhouse gas emissions are the basis for the quantification of greenhouse gas emissions from individual countries and the elaboration of country level mitigation strategies. For instance, the IPCC and the Food and Agriculture Organization of the United Nations (FAO) have developed two methodologies (IPCC, 2006; FAO, 2010) to estimate methane emissions from cattle enteric fermentation. Both methodologies rely on the use of methane emissions factors (Y_m) which estimate emissions as the proportion of the animal's gross energy intake which is lost as methane. Nevertheless, the Y_m introduced by the IPCC do not directly represent variations in methane emissions determined by the ruminal fermentation of distinct carbohydrates. More specifically, the ruminal fermentation of distinct carbohydrate fractions in the ration leads to the production of different volatile fatty acids (VFA), with a proportional increase in acetate production, and consequently methane, when structural carbohydrates are fermented (Moe & Tyrrell, 1979). In this context, the usefulness of Y_m based models in examining dietary related mitigation strategies is limited.

Moreover, the low predictive ability of the Y_m approach may introduce substantial error into the calculation of methane emissions in national and global greenhouse gas inventories (Ellis *et al.*, 2010). In a recent study, the prediction ability of several models to estimate methane emissions from enteric fermentation was evaluated and, in general, predictions were poor (Ellis *et al.*, 2010). The poor predictive ability of current models is mainly a result of the relatively small data sets used for model parameterization and the modeling techniques used. Furthermore, methane prediction models, to be used in determining national inventories and evaluating methane mitigation strategies, should not require excessive information as model inputs. For example, mechanistic models of rumen fermentation (Baldwin *et al.*, 1987; Dijkstra *et al.*, 1992) require numerous inputs, which are usually not available at the production system level. Therefore, a balance between model complexity and predictive ability must be achieved for the identification of the model with the best predictive ability from the set of models based on variables available at the production system level. Such prediction models should be preferably developed from databases which are the basis of the current energy evaluation systems (NRC, 2001) because the determination of animal energy requirements plays a major role in the estimation of emissions in national inventories.

In this context, the study objectives were to (i) identify key explanatory variables in the prediction of methane emissions; (ii) develop methane emission prediction equations using a large database of dairy and beef cattle; (iii) conduct a cross-validation of prediction equations; (iv) examine the improvement in predictive ability of methane equations with an increase in model complexity; and (v) investigate the appropriateness of current models used in determining methane emissions from cattle enteric fermentation.

Material and methods

Database

A database containing 2574 indirect respiration calorimetry records of dairy and beef cattle in 62 studies conducted from 1963 to 1995 in the former USDA Energy Metabolism Unit at Beltsville, Maryland was assembled by Wilkerson *et al.* (1995, 1997). Indirect respiration calorimetry is the gold standard for measuring energy flows between the animal and its environment. The database is a reliable representation of the current state of knowledge on methane emissions from cattle enteric fermentation and is the basis of the current energy evaluation system for feeding dairy cattle (NRC, 2001). The database comprises individual records of total energy balance trials from Holstein and Jersey lactating ($n = 1,111$) and nonlactating ($n = 591$) dairy cows; Holstein, Jersey and Angus-Hereford cross heifers ($n = 414$) and Holstein, Angus, Hereford and Angus-Hereford cross steers ($n = 458$). A general description of the database and summary statistics of nutrient composition of diets and animal information are available as supporting information (Tables S1 and S2). A comprehensive description of methane measurement techniques is available from Moe & Tyrrell (1979). The database was divided into four sub-data sets: Lactating cows, nonlactating cows, heifers and steers for which the data analysis is described below.

Model selection

Model development was conducted in a sequential approach, with increasing model complexity at each level. The objective of this sequential approach was similar to the IPCC (2006), which is based on different levels of available information. Three complexity levels were specified, namely gross energy, dietary, and animal levels. In the gross energy level, emissions are predicted using the animal's gross energy intake. In the dietary complexity level, diet characteristics (fiber fractions, crude protein, ether extract, and metabolizable energy), as well as gross energy intake, are potential predictors. In the animal complexity level, milk composition (fat, protein, and nonfat soluble), and animal information (body weight, and breed), as well as variables from the dietary level are potential predictors. Gross energy intake was selected as a measure of animal's feed consumption to be consistent and comparable with current national greenhouse gas inventories and to examine methane emissions from an energy loss perspective.

However, this explanatory variable may not be available to some model users. Therefore, three equations are provided as supporting information for the estimation of dietary gross energy based on diet nutrient composition (Eqns S1–S3). Along with dry matter intake, such equations may be accurately used in the calculation of the animal gross energy intake (see Fig. S1). Explanatory variables, which play a key role in predicting emissions, at the dietary and animal complexity levels were selected through a Bayesian model selection procedure, namely a Reversible Jump Markov Chain Monte Carlo (RJMCMC) sampler (Lunn *et al.*, 2009). The RJMCMC is a natural extension of the Metropolis-Hastings algorithm, in which the sampler is allowed to ‘jump’ between models with parameter space of different dimensions. Model selection is then based on the competing models posterior probabilities (Lunn *et al.*, 2009). Variables were selected under a linear hierarchical model, which is described in the following section. The RJMCMC sampler was implemented in the Bayesian statistical software WinBUGS (Lunn *et al.*, 2000). The convergence of the RJMCMC was assessed by the specification of two chains with over-dispersed initial states, the monitoring of the trace plot of model state variable and the assessment of similarities between posterior model probabilities generated by each chain, as described in Lunn *et al.* (2009).

Model specification and implementation

In the second part of model development, parameters were estimated under the selected most probable models at each level of complexity. A Bayesian framework was adopted, in which parameters were estimated by Markov Chain Monte Carlo methods (MCMC) in the statistical software WinBUGS (Lunn *et al.*, 2000). The Bayesian framework is particularly interesting here because it naturally accommodates the hierarchical structure of the data. Moreover, it allows simple fitting of models with residuals modeled through a Student’s *t* distribution, providing robustness for the presence of outlying and extreme observations (Gelman *et al.*, 2004). Two chains with over-dispersed initial values were specified for each parameter and chain mixing, auto-correlation, posterior densities and the Gelman-Rubin diagnostics (Gelman & Rubin, 1992) were used to visually assess chain convergence and determine the required burn in period. The tests of the convergence diagnostic and output analysis (CODA) package (Best *et al.*, 1995) were used to formally assess chains’ convergence. After discarding the burn-in samples, the sampler was run for additional 250 000 iterations with chain thinning of 10, generating a posterior of 50 000 samples based on two chains. The linear hierarchical model is described in terms of three distinct hierarchy stages, in which tilde (~) is used to indicate ‘distributed as’ and bold font is used to indicate vector or matrix.

First stage hierarchy. The first stage specifies a model for the data, conditional on model parameters, through a Student’s *t* distribution: $y_{ijk} | \beta, \alpha_i, \delta_j, \sigma_e^2, v_e \sim t(x_{ijk}^T \beta + \alpha_i + \delta_j, \sigma_e^2, v_e)$, where y_{ijk} denotes the k^{th} record ($k = 1, \dots, n_{ij}$) on the i^{th} animal ($i = 1, \dots, I$) in the j^{th} study ($j = 1, \dots, J$), β is the vector of ‘fixed’ (in the frequentist sense) regression coefficients, α_i is

the ‘random’ effect of the i^{th} animal, δ_j is the ‘random’ effect of the j^{th} study, σ_e^2 and v_e are the scale parameter and the degrees of freedom parameter of the Student’s *t* distribution. Further, superscript T denotes transpose and x_{ijk} is the known vector of covariates selected by the RJMCMC sampler (Note that the distinction between ‘fixed’ and ‘random’ effects does not arise in the Bayesian setting since all unknowns are treated as random variables. In this manuscript, ‘fixed’ effects have minimally informative priors with known hyperparameters and ‘random’ effects have priors for which hyperparameters remain to be estimated).

Second stage hierarchy. The second stage specifies distributions for the ‘random’ effects conditional on their variance components: $\alpha_i | \sigma_a^2 \sim N(0, \sigma_a^2)$; $\delta_j | \sigma_s^2 \sim N(0, \sigma_s^2)$, where N denotes the normal distribution, σ_a^2 and σ_s^2 are the animal and study variance components.

Third stage hierarchy. The third stage specifies prior distributions for model parameters: $\beta \sim N_p(\mathbf{0}, 10^6 \mathbf{I}_p)$; $\sigma_e^{-2} \sim \text{Gam}(10^{-3}, 10^{-3})$; $v_e \sim U(2, 100)$; $\sigma_a \sim U(0, 100)$; $\sigma_s \sim U(0, 100)$ where N_p denotes the p -dimensional multivariate normal distribution, \mathbf{I}_p is the p -dimensional identity matrix, Gam denotes the gamma distribution (for which $\text{Gam}(a, b)$ has mean = a/b and variance = a/b^2) and U denotes the uniform distribution. Gamma and uniform distributions are standard choices of prior distributions to model variance components (Gelman *et al.*, 2004). Furthermore, this choice of prior distributions is based on the construction of minimally informative priors (Gelman *et al.*, 2004) while satisfying the conditions that variance components are positive and the degrees of freedom parameter of the Student’s *t* distribution is greater than two.

Cross-validation and model evaluation

The predictive ability of fitted methane prediction models was evaluated through K-fold cross-validation (Efron & Tibshirani, 1993) in which folds were composed of individual studies. In short, the cross-validation scheme generates prediction of methane emissions of each fold through a model fitted without observations from that fold. The ability of the fitted model in predicting out of sample observations is then evaluated by the mean square prediction error (MSPE) as described in Bibby & Toutenburg (1977). An estimate of the model MSPE is derived by averaging the MSPE across folds with the number of observations from each fold used as a weighting factor. The root mean square prediction errors (RMSPE), that is, the square root of the MSPE’s, are reported in two distinct forms. In the first form, RMSPE’s are reported in the same units which the methane observations were recorded. This form of the RMSPE can be used as a measure of the uncertainty associated with the prediction equations in the sense that it is an estimate of the mean prediction error. In the second form, RMSPE’s are reported as a proportion of methane emission means (observed methane emission means ranged 6.06 to 16.57 MJ/d across the four data sets – see supporting information). This second form has the advantage of enabling the comparison of the RMSPE from different studies and can be

used to compare the predictive ability of models fitted with different data sets.

In addition, the stability of fitted models was examined through multicollinearity diagnostics, namely variance inflation factors, condition number and the determinant of the predictor's correlation matrix (Belsley *et al.*, 1980). Furthermore, the animal intra-class correlation, or the repeatability of methane emissions, was estimated as the ratio of the animal variance component and the total variance, that is $r = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2 + \sigma_v^2 / (v_e - 2)}$. Fitted prediction models, for each data set, were compared among complexity levels using the Deviance Information Criteria (DIC). The DIC is a model assessment tool to examine the trade-off between model goodness of fit and complexity, with differences of 5 and 10 DIC units showing a tendency and a substantive improvement respectively (Spiegelhalter *et al.*, 2002). Predictions of methane emissions using literature models were: IPCC [CH_4 (MJ/d) = 0.065 × GEI (MJ/d)] and FAO [CH_4 (MJ/d) = {9.75 - 0.05 × Dry Matter digestibility, (%) / 100} × GEI (MJ/d)].

Results

The most probable models for each complexity level and data set as well as their posterior probabilities are presented in Table 1. Models' posterior probabilities were calculated as the proportion of chain iterations the RJMCMC sampler spent in each competing model and

were used to identify the most probable model conditional on the observed data. Across all complexity levels and data sets, a most probable model was identified with relatively large posterior probability ($P > 0.64$). Gross energy intake was the key variable in predicting methane emissions and was present in the selected models across all complexity levels and data sets. The variable's marginal posterior probability (as defined in Lunn *et al.*, 2009) of gross energy intake being included in the selected models was equal to unity in all four data sets. Neutral detergent fiber (NDF) and ether extract (EE) dietary proportions represent the fiber and lipid contents of the diet and were the only diet descriptors which were selected by the RJMCMC sampler. More specifically, dietary NDF proportion was included in the most probable model of the lactating cows' and heifers' data set and dietary EE proportion was included in the lactating and nonlactating cows' data sets.

When the RJMCMC algorithm was allowed to sample from models for which the parameter space contained animal information and milk composition as possible covariates, body weight was selected in three data sets. More specifically, body weight marginal posterior probabilities were large ($P > 0.98$) in the lactating cows, heifers and steers datasets. The marginal

Table 1 Model posterior probabilities for the dietary and animal complexity levels and prediction equations for each complexity level [Gross energy (GE), Animal and Dietary] and data set (Lactating cows, Nonlactating cows, Heifers, and Steers)

Model	Posterior Probability*	Prediction Equation†
LACTATING COWS		
GE Level	–	$\text{CH}_4 = 3.247 (0.429) + 0.043 (0.001) \times \text{GEI}$
Dietary Level	0.74	$\text{CH}_4 = 0.225 (0.713) + 0.042 (0.001) \times \text{GEI} + 0.125 (0.015) \times \text{NDF} - 0.329 (0.094) \times \text{EE}$
Animal Level	0.86	$\text{CH}_4 = -9.311 (1.060) + 0.042 (0.001) \times \text{GEI} + 0.094 (0.014) \times \text{NDF} - 0.381 (0.092) \times \text{EE} + 0.008 (0.001) \times \text{BW} + 1.621 (0.119) \times \text{MF}$
NONLACTATING COWS		
GE Level	–	$\text{CH}_4 = 2.381 (0.153) + 0.053 (0.001) \times \text{GEI}$
Dietary Level	0.85	$\text{CH}_4 = 2.880 (0.200) + 0.053 (0.001) \times \text{GEI} - 0.190 (0.049) \times \text{EE}$
Animal Level	0.64	$\text{CH}_4 = 2.880 (0.200) + 0.053 (0.001) \times \text{GEI} - 0.190 (0.049) \times \text{EE}$
HEIFERS		
GE Level	–	$\text{CH}_4 = 1.289 (0.185) + 0.051 (0.001) \times \text{GEI}$
Dietary Level	0.84	$\text{CH}_4 = -0.163 (0.298) + 0.051 (0.001) \times \text{GEI} + 0.038 (0.006) \times \text{NDF}$
Animal Level	0.86	$\text{CH}_4 = -1.487 (0.318) + 0.046 (0.001) \times \text{GEI} + 0.032 (0.005) \times \text{NDF} + 0.006 (0.0007) \times \text{BW}$
STEERS		
GE Level	–	$\text{CH}_4 = 0.743 (0.119) + 0.054 (0.001) \times \text{GEI}$
Dietary Level	0.83	$\text{CH}_4 = 0.743 (0.119) + 0.054 (0.001) \times \text{GEI}$
Animal Level	0.87	$\text{CH}_4 = -0.221 (0.151) + 0.048 (0.001) \times \text{GEI} + 0.005 (0.0005) \times \text{BW}$

*The posterior probability of a competing model m , conditional on the observed data is estimated by the proportion of chain iterations the reversible jump sampler spent in model m .

†Equations are presented as parameter posterior means and standard deviation in parenthesis. CH_4 , Methane emissions (MJ/d); GEI, Gross energy intake (MJ/d); NDF, Dietary neutral detergent fiber proportion (% of dry matter); EE, Dietary ether extract proportion (% of dry matter); BW, Body Weight (kg); MF, Milk fat (%).

probability of body weight inclusion was 0.12 in the steers' dataset. Milk fat percentage was also present in the most probable model at the animal complexity level for the lactating cows' data set. The high marginal posterior probability of milk fat being included in the selected models ($P = 1$) suggests that milk fat percentage has an important role in predicting methane emissions from enteric fermentation. Cattle's breed marginal posterior probabilities were very low for all data sets ($P < 0.13$). The effect of breed on methane emissions was not evident on any animal category, suggesting that there are no differences on methane emissions between the examined breeds when dietary nutrient composition, gross energy intake and body weight are already accounted in the model.

Estimates of variance components, repeatability, the degrees of freedom parameter of the Student's t distribution as well as their 95% Monte Carlo Credible Intervals are in Table 2. Model evaluation statistics are in Table 3. The increase in model complexity, that is, allowing the RJMCMC to sample from models which contain dietary and animal information as covariates, resulted in models which are better supported by the observed data. For all data sets, models fitted at the dietary and animal complexity levels presented smaller DIC's (Table 3) than models fitted with gross energy intake as the only predictor. Moreover, models fitted at the animal complexity level presented smaller DIC's

than models fitted at the dietary complexity level, given that different models were selected at these two complexity levels. Therefore, including dietary and animal information improved model goodness of fit and resulted in models which are better supported by the observed data. The RMSPE's from fitted models and from the IPCC (2006) and FAO (2010) models are reported in Table 3. From a prediction perspective, the inclusion of the selected dietary variables as predictors did not reduce prediction errors substantially when compared with models fitted with gross energy intake as the only predictor. Conversely, the inclusion of animal information, more specifically body weight and milk fat percentage in the steers and lactating cow's data sets, respectively, reduced substantially the RMSPE when compared with models from the gross energy intake and dietary complexity levels.

As a final step of model evaluation, we have calculated variance inflation factors, condition numbers and the determinant of the predictor's correlation matrix for models fitted at the dietary and animal complexity levels (Table 4). These calculations investigate the degree of multicollinearity in the fitted models because it can be expected, from a biological perspective, that some covariates in the selected models might be moderately correlated. For example, gross energy intake with body weight and dietary NDF proportion with milk fat percentage. Variance inflation factors were below 1.5 for

Table 2 Variance components, repeatability and degrees of freedom parameter of the Student's t distribution of the models from the gross energy (GE), Animal and Dietary complexity levels in each data set (Lactating cows, Nonlactating Cows, Heifers, and Steers)*, †.

Model	σ_e^2	σ_a^2	σ_s^2	r	v_e
LACT. COWS					
GE Level	2.69 (2.17, 3.30)	2.38 (1.72, 3.17)	3.09 (1.75, 5.18)	0.23 (0.16, 0.31)	4.62 (3.26, 6.78)
Dietary Level	2.42 (1.96, 2.97)	2.46 (1.79, 3.26)	4.04 (2.31, 6.83)	0.22 (0.16, 0.30)	4.60 (3.21, 6.87)
Animal Level	2.49 (2.02, 3.02)	1.74 (1.20, 2.39)	2.50 (1.31, 4.47)	0.22 (0.15, 0.30)	7.14 (4.28, 13.03)
NONLACTATING					
GE Level	0.29 (0.22, 0.37)	0.57 (0.38, 0.80)	0.81 (0.45, 1.31)	0.26 (0.17, 0.37)	3.38 (2.44, 4.71)
Dietary Level	0.28 (0.21, 0.35)	0.56 (0.38, 0.79)	0.87 (0.48, 1.49)	0.28 (0.21, 0.35)	3.28 (2.40, 4.52)
Animal Level	0.28 (0.21, 0.35)	0.56 (0.38, 0.79)	0.87 (0.48, 1.49)	0.28 (0.21, 0.35)	3.28 (2.40, 4.52)
HEIFERS					
GE Level	0.50 (0.36, 0.68)	0.15 (0.07, 0.27)	0.59 (0.17, 1.63)	0.08 (0.03, 0.16)	4.11 (2.56, 7.01)
Dietary Level	0.48 (0.35, 0.66)	0.12 (0.04, 0.22)	1.33 (0.42, 3.66)	0.05 (0.02, 0.11)	4.67 (2.68, 9.11)
Animal Level	0.37 (0.26, 0.51)	0.10 (0.04, 0.19)	0.89 (0.27, 2.44)	0.06 (0.02, 0.12)	3.73 (2.36, 6.25)
STEERS					
GE Level	0.21 (0.15, 0.28)	0.09 (0.04, 0.15)	0.63 (0.18, 1.93)	0.04 (0.005, 0.10)	2.45 (2.02, 3.25)
Dietary Level	0.21 (0.15, 0.28)	0.09 (0.04, 0.15)	0.63 (0.18, 1.93)	0.04 (0.005, 0.10)	2.45 (2.02, 3.25)
Animal Level	0.17 (0.13, 0.23)	0.09 (0.04, 0.15)	0.35 (0.09, 1.05)	0.07 (0.01, 0.14)	2.61 (2.04, 3.55)

*Parameters are reported as posterior means and 95% Monte Carlo Credible Intervals in parentheses.

† σ_e^2 is the scale parameter of the Student's t distribution, σ_a^2 is the animal variance component, σ_s^2 is the study variance component, r is the repeatability or the intraclass correlation {i.e., $\sigma_a^2/(\sigma_a^2 + \sigma_s^2 + \sigma_e^2 v_e/(v_e - 2))$ } and v_e is the degrees of freedom parameter of the Student's t distribution.

Table 3 Deviance information criteria and root mean square prediction error (MJ/d and% of the CH₄ mean) for the fitted equations in each complexity level [Gross energy (GE), Animal and Dietary] for each data set (Lactating cows, nonlactating cows, heifers, and steers) and for the IPCC (2006) and FAO (2010) models

Model	DIC*	RMSPE (MJ/d)†	RMSPE (% of CH ₄ mean)†
LACTATING COWS			
GE Level	1.811	3.01	18.14
Dietary Level	1.707	2.97	17.89
Animal Level	1.402	2.59	15.62
IPCC (2006)	–	5.06	30.50
FAO (2010)	–	4.93	29.74
NONLACTATING COWS			
GE Level	–225.6	1.25	13.95
Dietary Level	–243.1	1.29	14.42
Animal Level	–243.1	1.29	14.42
IPCC (2006)	–	1.67	18.61
FAO (2010)	–	1.92	21.40
HEIFERS			
GE Level	–19.97	1.21	18.25
Dietary Level	–56.02	1.25	18.88
Animal Level	–119.9	1.23	18.55
IPCC (2006)	–	1.23	18.55
FAO (2010)	–	1.27	19.15
STEERS			
GE Level	–253.3	1.00	16.47
Dietary Level	–253.3	1.00	16.47
Animal Level	–351.0	0.92	15.10
IPCC (2006)	–	1.17	19.22
FAO (2010)	–	1.29	21.28

*Deviance information criteria. Trade-off between model goodness-of-fit and Complexity (Spiegelhalter *et al.*, 2002). Models with smaller DIC are better supported by the data.

†Square root of the mean square prediction error, expressed in MJ/d and as a percentage of the methane emissions means.

all models. Multicollinearity issues often arise when variance inflation factors are larger than 10 (Chatterjee *et al.*, 2000). The determinant of the predictor's correlation matrix ranges from zero to one, in which a model with a perfect linear combination between predictors has a determinant of zero and a model with no multicollinearity has a determinant of one. The determinants, in our models, were substantially distant from zero, ranging from 0.53 to 0.99. Finally, the condition number indicates possible issues with multicollinearity when the number is between 10 and 30 and the presence of multicollinearity when the number is greater than 30 (Belsley *et al.*, 1980). The condition numbers from our study are below 1.95, confirming that multicollinearity is not an issue with our prediction models and parameters are stable.

Table 4 Variance inflation factors, determinant of the predictor's correlation matrix and condition numbers for the dietary and animal complexity levels models in each data set (Lactating cows, nonlactating cows, heifers, and steers)

Model	Largest VIF*	R †	Condition Number‡
Lactating Cows			
Dietary Level	1.23	0.81	1.59
Animal Level	1.46	0.53	1.95
NonLactating Cows			
Dietary Level	1.09	0.99	1.10
Animal Level	1.09	0.99	1.10
Heifers			
Dietary Level	1.02	0.98	1.15
Animal Level	1.25	0.79	1.64
Steers			
Dietary Level	–	–	–
Animal Level	1.28	0.78	1.67

*Largest variance inflation factor for each model. VIF's = $\text{diag}\{\mathbf{R}^{-1}\}$, where \mathbf{R} is the predictors' correlation matrix. Multicollinearity issues when VIF > 10.

†Determinant of the predictor's correlation matrix \mathbf{R} , 0 indicates 100% collinearity, 1 indicates no collinearity.

‡Condition Number = $\sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$ where λ denotes eigenvalue of \mathbf{R} , >30 indicates multicollinearity issues.

Discussion

Key explanatory variables to predict methane emissions were identified for various animal categories and used to fit linear hierarchical models. Dietary NDF and EE proportions were identified as key dietary variables in predicting enteric methane emissions of distinct cattle categories. The type of dietary carbohydrates (structural vs. nonstructural) fermented in the rumen plays a major role in determining the profile of ruminal VFA's (Murphy *et al.*, 1982) and consequently on methane emissions (Moe & Tyrrell, 1979). The absence of NDF in the selected models under the nonlactating cows' and steers' data sets suggests that factors other than dietary carbohydrate type have a major role in determining methane emissions in these two data sets. In the study by Moe & Tyrrell (1979), the effect of carbohydrate type on methane emissions was relatively less important at low intake levels, suggesting that methane emissions of cattle at low intake levels can be predicted from dry matter or digestible carbohydrate intake. Furthermore, the slopes of the NDF variable were positive (and 'significantly' different from zero, in the sense that their 95% Monte Carlo Credible Intervals did not overlap zero) across all complexity levels and selected data sets. Therefore, methane emissions are on average reduced when comparing pairs of diets that have reduced NDF proportion and are identical in the other covariates.

The increase in methane emissions with increased dietary fiber proportion is consistent with results from the literature because the role of dietary fiber proportion and carbohydrate type on ruminal fermentation and methane emissions is well established (Moe & Tyrrell, 1979; Murphy *et al.*, 1982).

Dietary EE proportion, which represents the lipid component of feedstuffs, was present in the models selected for the lactating cows and nonlactating cows' data sets. The inclusion of lipids in the diet has been reported by recent studies (Martin *et al.*, 2010; Grainger & Beauchemin, 2011) as potential methane mitigation strategies. However, the steers and heifers most probable models did not contain dietary ether extract. Dietary EE marginal posterior probabilities were 0.07 and 0.003 for the heifers' and steers' data sets, respectively. Nevertheless, it is important to note that not only the proportion of dietary lipid but also the lipid form and type, has the potential of altering ruminal fermentation and consequently methane emissions. The efficacy of different forms of dietary lipid supplementation is not addressed by our analysis and can be found elsewhere (Martin *et al.*, 2010; Grainger & Beauchemin, 2011). The slope of the EE covariate was negative (and 'significantly' different from zero, in the sense that their 95% Monte Carlo Credible Intervals did not overlap zero) for all complexity levels and selected data sets, indicating that methane emissions are decreased when dietary EE content is increased. The magnitude of the emissions response to dietary fat in our study ranged from 0.19 to 0.38 (MJ of methane/% of dietary EE), suggesting that there is an average decrease of 0.19–0.38 MJ of methane for every percentage increase in dietary EE when other covariates are at identical level.

Similarly, animals' body weight was selected by the RJMCMC algorithm as a key explanatory variable in predicting emissions. The importance of body weight in explaining methane emissions from enteric fermentation can be attributed to the relationship between body weight and gut capacity, because gut volume is proportional to body weight (Demment & Van Soest, 1985). More specifically, Smith & Baldwin (1974) found that ruminal weight was proportional to body weight of dairy cows. Furthermore, the kinetics of ruminal feed particles plays a key role in diet digestibility which is dependent on rumen volume and feed intake (Van Soest, 1994). In this context, animals with different body weights, when consuming similar amounts of a common diet, would present different ruminal passage rates and amounts of fermented feed. This difference in ruminal kinetics would alter volatile fatty acid production and consequently methane emissions. Likewise, milk fat percentage was also included in the most probable model at the animal complexity level for the

lactating cows' data set. Fat is the most variable component in milk and is highly dependent on diet composition and ruminal fermentation characteristics (Bauman & Griinari, 2003). The positive coefficient of the milk fat percentage may be explained by the relationship between acetate production and milk fat content. Acetate is the major precursor for de novo milk fat synthesis and acetate production in the rumen is associated with a net generation of hydrogen which is available for methanogenesis. Moreover, recent studies have determined relationships between milk fatty acid composition and methane emissions and concluded that milk fatty acid composition may be potential predictors of rumen fermentation and methane emissions when used along with dietary variables (Dijkstra *et al.*, 2011; Mohammed *et al.*, 2011). Finally, the effect of breed on methane emissions was not evident on any animal category, suggesting that there are no differences on methane emissions between the examined breeds when dietary nutrient composition, gross energy intake and body weight are already accounted in the model. The absence of breed based differences on methane emissions when feed intake and production status are controlled are consistent with results from the literature (Munger & Kreuzer, 2006, 2008; Klevenhusen *et al.*, 2011) in which no differences between breeds of dairy cattle on methane emissions were observed.

It is beyond the scope of our study to conduct a complete analysis of methane emissions genetic parameters. However, from an animal breeding perspective, the intraclass correlation represents the repeatability, which is the proportion of the phenotypic variance due to genetic and permanent environmental effects and sets the upper bound of the heritability (Falconer & Mackay, 1996). Recently, there has been an increasing interest in estimates of methane emissions repeatabilities and heritabilities in ruminants because breeding for low emitting animals has been suggested as a long term mitigation strategy (De Haas *et al.*, 2011). Our repeatabilities estimates were substantially lower than the ones from the literature. In the lactating and nonlactating cows' data sets, repeatabilities ranged from 0.22 to 0.28 and in the heifers and steers data sets, repeatabilities ranged from 0.04 to 0.08. Repeatability estimates of 0.59 for daily methane emissions were reported by Vlaming *et al.* (2008) in nonlactating dairy cows in New Zealand. It is important to note that although dairy cows in New Zealand may be from similar lineage to northern American Holstein cows, they have been selected for different traits over the past five decades (milk production on grazing systems vs. milk production on relatively higher concentrate diets). Similarly, De Haas *et al.* (2011) estimated heritabilities of predicted daily methane emissions of 0.35 for dairy cows in the Netherlands.

It is important to note that our relatively low repeatability estimates may be a result of the cross-sectional structure of the data. Most of the studies were conducted under Latin square and crossover experimental designs in which animals received distinct dietary treatments at each experimental period. It is possible that higher repeatability estimates would be derived if methane measurements were taken throughout the lactation or feeding period instead of under distinct dietary experimental treatments.

Generally, including dietary and animal information improved model goodness-of-fit and resulted in models which are better supported by the observed data. The identification of models which are better supported by the data plays an important role in the examination of methane emissions mitigation strategies and environmental policies. The misspecification of the model structure can introduce bias in parameter estimation and the conclusions about the response of methane emissions to changes in explanatory variables are likely to be incorrect. For example, examining the effects of dietary fat proportion on methane emissions from a model with dietary fat as the only predictor can cause the estimation of biased parameters and incorrect conclusions if other parameters that influence the response are not accounted in the model. Further, the RMSPE's of the estimated prediction models were substantially lower than the ones from the IPCC (2006) and FAO (2010) prediction models, either at the gross energy, dietary or animal complexity level. For instance, in the lactating cows' data set, the model fitted with the gross energy intake as its only covariate has a substantially lower prediction error than the IPCC model (18.14 vs. 30.50%). Across all data sets, the models developed only using gross energy intake as predictors have a lower RMSPE than the models from the IPCC (2006) and FAO (2010). Therefore, using the models fitted in this study, methane emissions from Northern American cattle can be predicted using the same amount of information required by the IPCC (2006) Tier II methodology, but with a substantial reduction in the prediction error. Moreover, RMSPE's from our methane prediction equations are smaller than the ones calculated by Ellis *et al.* (2010) in which several methane prediction models were evaluated by a literature derived database and a database of individual cows. The magnitude of the RMSPE of our prediction equations is also smaller than the ones calculated by Ellis *et al.* (2007) and Mills *et al.* (2003) in which linear and nonlinear models were developed and evaluated using a literature database and data from the Metabolism Unit of the Centre for Dairy Research at the University of Reading, respectively.

It is important to note that prediction models have intrinsic limitations. For example, most diets used in

model development had relatively high-forage contents compared to a typical northern American beef cattle feedlot diets. Consequently, equations for steers and heifers may not be applicable to finishing feedlot diets. Furthermore, data were collected in respiration chambers and may not be applicable to grazing conditions. Dietary nutrient composition may be considerably different in grazing systems and feeding behavior may also affect emissions. Therefore, equations should be used with caution if predictions are to be made in grazing systems. With such limitations in mind, we propose various models for predicting emissions from dairy and beef cattle which require different levels of information and can be used in the development of greenhouse gas inventories having various levels of detail. In particular, prediction models with gross energy intake as the only explanatory variable outperformed the IPCC tier II methodology and can be readily implemented in greenhouse gas inventories because both methodologies use the same explanatory variable. Further, the use of the models developed in this study will advance the examination of diet related methane emissions mitigation strategies and regulatory policies.

Acknowledgements

Research was partially supported by the UC Davis Sesnon Endowed Chair program and the W. K. Kellogg Endowment, USDA NIFA Multistate Research Project NC-1040. We gratefully acknowledge the infrastructure support of the Department of Animal Science, College of Agricultural and Environmental Sciences, and the California Agricultural Experiment Station of the University of California, Davis.

References

- Baldwin RL, France J, Gill M (1987) Metabolism of the lactating cow I. Animal elements of a mechanistic model. *Journal of Dairy Research*, **54**, 77–105.
- Bauman DE, Griinari JM (2003) Nutritional regulation of milk fat synthesis. *Annual Review of Nutrition*, **23**, 203–227.
- Belsley DA, Kuh E, Welsch RE (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, New York, NY.
- Best N, Cowles MK, Vines SK (1995) *CODA Manual Version 0.30*. MRC Biostatistics Unit, Cambridge, UK.
- Bibby J, Toutenburg T (1977) *Prediction and Improved Estimation in Linear Models*. John Wiley & Sons, Chichester.
- Chatterjee S, Hadi AS, Price B (2000) *Regression Analysis by Example*. 3rd ed. John Wiley & Sons, Inc., New York, NY.
- De Haas Y, Windig JJ, Calus MPL, Dijkstra J, de Haan M, Bannink A, Veerkamp RF (2011) Genetic parameters for predicted methane production and potential for reducing enteric emissions through genomic selection. *Journal of Dairy Science*, **94**, 6122–6134.
- Demment MW, Van Soest PJ (1985) A nutritional explanation for body-size patterns of ruminant and nonruminant herbivores. *The American Naturalist*, **125**, 641–672.
- Dijkstra J, Neal HD, Beaver DE, France J (1992) Simulation of nutrient digestion, absorption and outflow in the rumen: model description. *Journal of Nutrition*, **122**, 2239–2256.
- Dijkstra J, Van Zijderveld SM, Apajalahti JA *et al.* (2011) Relationships between methane production and milk fatty acid profiles in dairy cattle. *Animal Feed Science and Technology*, **166–167**, 590–595.

- Efron B, Tibshirani RJ (1993) *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- Ellis JL, Kebreab E, Odongo NE, McBride BW, Okine EK, France J (2007) Prediction of methane production from dairy and beef cattle. *Journal of Dairy Science*, **90**, 3456–3467.
- Ellis JL, Bannink A, France J, Kebreab E, Dijkstra J (2010) Evaluation of enteric methane prediction equations for dairy cows used in whole farm models. *Global Change Biology*, **16**, 3246–3256.
- EPA (2011) *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2009*. EPA 430-R-11-005. Environmental Protection Agency, Washington, DC.
- Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*. 4th ed. Longman, London.
- FAO (2010) *Greenhouse Gas Emissions from the Dairy Sector*. Food and Agriculture Organization of the United Nations, Rome, Italy.
- FAO (2011) *The State of Food and Agriculture*. Food and Agriculture Organization of the United Nations, Rome, Italy.
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–472.
- Gelman A, Carlin J, Stern H, Rubin D (2004) *Bayesian data analysis*. 2nd ed. Texts in Statistical Science. Chapman and Hall, London, UK.
- Grainger C, Beauchemin KA (2011) Can enteric methane emissions from ruminants be lowered without lowering their production? *Animal Feed Science and Technology*, **166–167**, 308–320.
- IPCC (2006) *IPCC Guidelines for National Greenhouse Gas Inventories*. Institute for Global Environmental Strategies, Hayama, Japan.
- Johnson KA, Johnson DE (1995) Methane emissions from cattle. *Journal of Animal Science*, **73**, 2483–2492.
- Klevenhusen F, Bernasconi SM, Kreuzer M, Soliva CR (2011) Experimental validation of the intergovernmental panel on climate change default values for ruminant-derived methane and its carbon-isotope signature. *Animal Production Science*, **50**, 159–167.
- Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–337.
- Lunn DJ, Best N, Whittaker JC (2009) Generic reversible jump MCMC using graphical models. *Statistics and Computing*, **19**, 395–408.
- Martin C, Morgavi DP, Doreau M (2010) Methane mitigation in ruminants: from microbe to the farm scale. *Animal*, **4**, 351–365.
- Mills JAN, Kebreab E, Yates CM *et al.* (2003) Alternative approaches to predicting methane emissions from dairy cows. *Journal of Animal Science*, **81**, 3141–3150.
- Moe PW, Tyrrell HF (1979) Methane production in dairy cows. *Journal of Dairy Science*, **62**, 1583–1586.
- Mohammed R, McGinn SM, Beauchemin KA (2011) Prediction of enteric methane output from milk fatty acid concentrations and rumen fermentation parameters in dairy cows fed sunflower, flax, or canola seeds. *Journal of Dairy Science*, **94**, 6057–6068.
- Moraes LE, Wilen JE, Robinson PH, Fadel JG (2012) A linear programming model to optimize diets in environmental policy scenarios. *Journal of Dairy Science*, **95**, 1267–1282.
- Munger A, Kreuzer M (2006) Methane emission as determined in contrasting dairy cattle breeds over the reproduction cycle. In: *Greenhouse Gases and Animal Agriculture: An update* (eds Soliva CR, Takahashi J, Kreuzer M), pp. 119–122. Elsevier, Amsterdam.
- Munger A, Kreuzer M (2008) Absence of persistent methane emission differences in three breeds of dairy cows. *Australian Journal of Experimental Agriculture*, **48**, 77–82.
- Murphy MR, Baldwin RL, Koong LJ (1982) Estimation of stoichiometric parameters for rumen fermentation of roughages and concentrate diets. *Journal of Animal Science*, **55**, 411–421.
- NRC (2001) *Nutrient Requirements of Dairy Cattle*. 7th ed. National Research Council, National Academy Press, Washington, DC.
- Oenema E (2004) Governmental policies and measures regulating nitrogen and phosphorus from animal manure in European agriculture. *Journal of Animal Science*, **82**, E196–E206.
- Smith NE, Baldwin RL (1974) Effects of breed, pregnancy, and lactation on weight of organs and tissues in dairy cattle. *Journal of Dairy Science*, **47**, 1055–1060.
- Spiegelhalter DJ, Best N, Carlin BP, Van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society: Series B*, **64**, 583–640.
- UN (2004) *World Population to 2300*. Population Division, United Nations, New York.
- Van Soest PJ (1994) *Nutritional Ecology of the Ruminant*. 2nd ed. Cornell University Press, Ithaca, NY.
- Vlaming JB, Lopez-Villalobos N, Brookes IM, Hoskin SO, Clark H (2008) Within and between animal variance in methane emissions in non-lactating dairy cows. *Australian Journal of Experimental Agriculture*, **48**, 124–127.
- Wilkerson VA, Casper DP, Mertens DR (1995) The prediction of methane production of Holstein cows by several equations. *Journal of Dairy Science*, **78**, 2402–2414.
- Wilkerson VA, Mertens DR, Casper DP (1997) Prediction of excretion of manure and nitrogen by Holstein dairy cattle. *Journal of Dairy Science*, **80**, 3193–3204.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Dietary composition, animal information, and references of the database individual studies.

Table S2. Dietary nutrient composition and animal status summary statistics of the four sub-data sets.

Equations S1–S3. Dietary gross energy prediction equations.

Figure S1. Diagnostic plots of gross energy intake prediction equations.