UC Davis
UC Davis Previously Published Works

Title

Predicting nitrogen excretion from cattle

Permalink

Journal

ISSN

Authors

Reed, KF
Moraes, LE
Casper, DP
et al.

Publication Date

DOI

Peer reviewed

Predicting nitrogen excretion from cattle

K. F. Reed,* L. E. Moraes,* D. P. Casper,† and E. Kebreab*[1]
*Department of Animal Science, University of California, Davis 95616
†Dairy Science Department, South Dakota State University, Brookings 57007

ABSTRACT

Manure nitrogen (N) from cattle production facilities can lead to negative environmental effects, such as contribution to greenhouse gas emissions, leaching and runoff to aqueous ecosystems leading to eutrophication, and acid rain. To mitigate these effects and to improve the efficiency of N use, accurate prediction of N excretion and secretions are required. A genetic algorithm was implemented to select models to predict fecal, urinary, and total manure N excretions, and milk N secretions from 3 classes of animals: lactating dairy cows, heifers and dry cows, and steers. Two tiers of model classes were developed for each category of animals based on model input requirements. A total of 6 models for heifers and dry cows and steers and an additional 2 models for lactating dairy cattle were developed. Evaluation of the models using K-fold cross validation based on all data and using the most recent 6 yr of data showed better prediction for total manure N and fecal N compared with urinary N excretion, which was the most variable response in the database. Compared with extant models from the literature, the models developed in this study resulted in a significant improvement in prediction error for fecal and urinary N excretions from lactating cows. For total manure production by lactating cows, extant and new models were comparable in their prediction ability. Both proposed and extant models performed better than the prediction methods used by the US Environmental Protection Agency for the national inventory of greenhouse gases. Therefore, the proposed models are recommended for use in estimation of manure N from various classes of animals.
Key words: modeling, climate change, livestock, nitrous oxide, manure

INTRODUCTION

The demand for animal products continues to grow and remains an essential component of sustainable food production. To meet the demand, the overall efficiency of milk and meat production needs to be increased without greater negative effects on the environment. Assessment of the efficiency of animal production and its subsequent environmental footprint requires quantification of each nutrient consumed, used, excreted, or lost to the environment. However, direct measurement of the required nutrient flows, in particular the quantity of manure nitrogen (N) excreted to the environment is challenging. Mathematical models are often used to estimate nutrient requirements, excretion, pollution, and greenhouse gas (**GHG**) emissions to quantify the efficiency and environmental effect of animal agriculture. National inventories of GHG emissions, including nitrous oxide (N_2O) emissions from livestock are calculated to assess and possibly mitigate the effect of emissions to global change (UN, 1992).

Accurate prediction of N_2O emissions and N load in fields and downstream ecosystems require that N excreted from individual animals or groups of animals be quantified with reasonable accuracy. Precise estimates of N excretion from livestock lead to better quantification of manure N, which is a basis for estimating N volatilization, leaching, run-off, and emission. Several models have been developed to predict N excretion from lactating dairy cattle, heifers, and steers to assess the efficiency of cattle production and calculate national inventories of N_2O emissions. For example, IPCC (2006) recommends methodologies for predicting N_2O emissions from livestock manure using either default (tier 1 or 2) or country specific estimates (tier 2 or 3) for N excretion. The accuracy of these estimates relies heavily on predicted N excretion because errors at the individual animal level are multiplied by the total national or regional animal population. The current IPCC Tier 2 default values for total N (**TN**) excretion are scalar multipliers of the total animal mass. In the US national GHG inventory, the US Environmental Protection Agency (**EPA**) estimates TN as the difference between total N intake (**NI**) and the sum of estimated secreted milk N (**MN**) and N retained in animal tissue. However, in addition to NI, diet composition and quality have been shown to be important predictors of N excretion in cattle and predict excretions with greater

accuracy when compared with models based on animal BW alone (Castillo et al., 2000; Nennich et al., 2005; Kebreab et al., 2010; Dijkstra et al., 2013). Moreover, the proportion of N excreted in urine (**UN**) and feces (**FN**) determines its fate as a source of N pollution (Tamminga, 1996); therefore, models predicting UN and FN separately improve estimation of nitrogenous compounds in air and soil. Independent UN and FN predictions could be used in process-based manure and soil models such as Manure-DNDC (Li et al., 2012) and result in more accurate environmental effect assessment.

Despite the number of studies dedicated to predicting N excretion from cattle, accuracy remains low (Higgs et al., 2012). Whereas the complexity of ruminant digestion plays a role, the current accuracy of N excretion models can also be attributed to statistical methodology, a narrow pool of potential explanatory variables, and small data sets for model development and selection. The objective of the current study was to develop mathematical models of N excretion with varying complexity for different classes of animals using a robust model selection process from a large data set and a wide pool of explanatory variables. We hypothesized that the resulting models would increase the accuracy of N excretion prediction models in cattle and subsequent estimates of N_2O emissions, thus leading to a better assessment of mitigation options to reduce GHG emissions.

MATERIALS AND METHODS

Database

The data set used for model development was collected from 59 indirect calorimetry studies conducted over 30 yr at the USDA Energy and Metabolism laboratory in Beltsville, Maryland. Indirect calorimetry is the preeminent method for measuring energy exchange between animals and their environment. Furthermore, the data provides a reliable representation of the current state of knowledge on nutrient utilization by lactating and growing cattle because of the number of observations over an extended period from the same research station and because it constitutes the basis of the nutrient evaluation system for the current US nutrient requirement models for dairy cattle (NRC, 2001). However, the data were collected in thermoneutral conditions fed temperate forages so care should be taken when extrapolating outside thermoneutral conditions. Studies for hormone therapy and those that used whole raw soybeans were excluded from the database because they were not representative of current US dairy practices. The database contains individual daily

nutrient balance observations from lactating Holstein and Jersey cows (n = 1,049); nonlactating Holstein and Jersey cows (n = 591); Holstein, Hereford, and Angus-Hereford cross heifers (n = 279); and Holstein, Angus, Hereford, and Angus-Hereford cross steers (n = 458). The data include a wide range of diet compositions that represents current diet formulations. Descriptions and summary statistics of relevant measured dietary and animal variables are provided in Table 1. Descriptions of the data collection and consolidation methods were described by Wilkerson et al. (1997). The data was divided by type of cattle into 3 sets for model development: lactating cows, nonlactating cows and heifers, and steers. Models were developed to predict feces, urine and total N for all 3 cattle types in addition to N in milk for lactating cattle.

Modeling Framework

The modeling framework in this study was composed of 3 complementary and sequential parts. In the first part, a linear mixed effects model was specified for the prediction of UN, FN, TN, and MN from the different animal classes. In the second part, key independent variables were identified using a multistage model-selection technique and the selected models were fitted to data. Finally, the fitted models were evaluated through a cross-validation procedure and compared with existing models from the literature. All models were fitted using the lme4 package in the R statistical software (Bates et al., 2013).

Model Specification

A cross-classified mixed-effects model was used to account for the dependence of records originating from the same animal and study. The model was specified as follows:

$$y_{ijk} = \mathbf{x}_{ijk}^{\mathrm{T}}\boldsymbol{\beta} + a_i + s_j + \varepsilon_{ijk},$$

where y_{ijk} is the kth observation of UN, FN, TN, or MN from the jth study and the ith animal; $\mathbf{x}_{ijk}^{\mathrm{T}}$ is the transpose vector of explanatory variables; $\boldsymbol{\beta}$ is the vector of fixed-effect parameters; a_i is the random effect associated with the ith animal; s_j is the random effect associated with the jth study; and ε_{ijk} is the residual random error assumed to be independent and identically distributed as $N(0, \sigma^2)$, where σ^2 is the residual variance. It was further assumed that a_i and s_j are independent and distributed as $a_i \sim N\left(0, \sigma_a^2\right)$ and $s_j \sim N\left(0, \sigma_s^2\right)$, where σ_a^2 is the between-animal variance and σ_s^2 is the between-study variance.

Table 1. Units, mean, SD, and range of variables used for covariate selection

Covariate[1]	Unit	Lactating cows (n = 1,047)				Heifers and dry cows (n = 870)				Steers (n = 458)			
		Mean	SD	Minimum	Maximum	Mean	SD	Minimum	Maximum	Mean	SD	Minimum	Maximum
UN	g/d	148	64.7	22.3	350	92.6	41.3	17.0	245	58.5	28.9	11.9	190
FN	g/d	144	47.4	35.1	334	50.1	18.7	11.8	125	46.6	21.1	11.0	117
TN	g/d	292	101	80.9	641	143	54.0	33.2	357	105	44.9	26.9	280
MN	g/d	112	43.7	0.85	267	—	—	—	—	—	—	—	—
Milk	kg/d	23.3	10.3	0.104	56.6	—	—	—	—	—	—	—	—
NI	g/d	432	145	79.9	932	156	57.8	39.2	352	130	53.7	43.4	307
DMI	kg/d	16.5	4.27	3.94	29.4	6.23	2.00	2.10	13.4	5.08	1.66	2.09	11.1
GE	MJ/kg of DM	18.9	0.52	17.8	20.7	19.0	0.887	16.0	23.0	19.4	1.26	17.3	23.4
DE	MJ/kg of DM	12.5	0.95	8.53	15.8	13.2	1.30	8.45	16.3	13.4	1.13	10.3	16.2
ME	MJ/kg of DM	10.8	0.90	6.85	14.6	11.0	1.19	6.51	14.3	11.5	1.04	8.16	14.1
OM	% of DM	93.6	1.11	87.9	96.3	93.0	2.27	77.9	96.9	93.8	1.78	86.4	96.8
DM	%	65.3	19.8	30.2	100	60.2	23.7	19.4	100	61.6	27.7	18.9	96.4
EE	% of DM	2.78	0.97	0.99	7.02	2.84	1.01	0.840	7.64	3.63	1.28	0.669	7.55
NDF	% of DM	34.3	7.44	14.9	76.1	38.9	12.9	13.2	78.3	35.8	14.7	18.7	74.7
ADF	% of DM	20.0	4.20	7.67	47.1	23.1	9.36	4.29	48.3	21.6	12.1	7.5	50.9
Lignin	% of DM	4.43	1.44	0.52	9.42	4.94	2.35	0.411	14.3	4.79	2.97	1.18	15.7
Ash	% of DM	6.39	1.11	3.71	12.13	6.97	2.21	3.08	22.1	6.25	1.78	3.22	13.6
CP	% of DM	16.2	2.51	5.17	23.5	15.6	2.64	4.92	23.0	15.9	3.56	10.8	25.4
Forage	% of DM	54.1	19.1	8.93	100	61.8	24.3	9.38	100	65.7	28.7	25.0	100
DIM	d	161	82	11	488	—	—	—	—	—	—	—	—
BW	kg	594	88.4	302	854	563	174	195	893	317	86.7	168	631

[1]UN = urinary nitrogen, FN = fecal nitrogen, TN = total manure nitrogen, MN = milk nitrogen, NI = N intake, GE = gross energy, DE = digestible energy, and EE = ether extract.

Model Selection

A multilevel, multistep selection process was implemented due to the large number of potential models. First, 2 levels of increasing model complexity were defined to provide flexibility for users with different levels of available input. The first model form (tier 1) was predicted from a single intake parameter: DMI (kg/d) or NI (g/d). The more complex models (tier 2) were selected from a large pool of diet and animal explanatory variables including dietary energy, protein, fiber, DM, fat, ash content, animal weight, breed, stage of lactation, and stage of pregnancy, where applicable (Table 1). All continuous variables were centered around their mean to reduce multicollinearity and facilitate computation.

A systematic procedure involving several steps to reduce the dimension of the variable pool was taken to decrease the number of potential models in tier 2. First, to prevent multicollinearity, one of a pair of highly correlated (r > 0.8) variables was taken out by removing the independent variable with smallest correlation with the response. Because this would remove all but one measure of intake and energy content, and because comparison between models with discrepant measures of these explanatory variables was desired, 6 sets of variable pools were developed for each model selection process in tier 2. These variable pools contained either NI or DMI and either gross energy, digestible energy (**DE**, kJ/kg of DM), or ME (kJ/kg of DM). In the second variable reduction step, variables with the greater degree of variance inflation were removed until the largest condition index of the independent variable correlation matrix was less than 10 as described by Belsley et al. (2004) using the perturb package in R statistical software (Hendrickx, 2012).

Having reduced multicollinearity to an acceptable degree, the final step used an information-theoretic approach to select a single model from all models within the domain defined by the established variable pools. The Bayesian information criterion (**BIC**) was chosen as the means for measuring model information because its theoretical objective is to select the model that best explains the underlying process leading to the observations and because of its tendency to select parsimonious models when the sample size is large (Ward, 2008).

A genetic algorithm was implemented to select the best models from each variable pool based on the BIC. Genetic algorithms, used in this case as an optimization tool, are a class of computer search and learning algorithms modeled after the biological process of genetic selection (Forrest, 1993). Heuristically, the genetic algorithm selects chromosomes, or models in this case,

to be passed on to future populations based on some measure of fitness (BIC). Chromosome information can be propagated through random mating with another chromosome or random mutation of a gene within a chromosome. The probability with which the algorithm populates the future generation through mating or mutation (p_m) and the probability of genetic mutation (p_g) can be tuned to facilitate thorough exploration of the parameter space and facilitate convergence. A key aspect of the algorithm is that the best chromosome from each generation is conserved without mutation, which guarantees nondecreasing fitness of each generation. In the present application, a chromosome was established as a binary vector of length v that indicates which variables are included in the design matrix **X**, where v is the number of independent variables in the largest possible model. The algorithm was initiated with 50 randomly selected chromosomes specifying an initial population of models that make up a random subset of models nested within the largest possible model. Future populations were propagated using the genetic algorithm with $p_m = 0.99$ and $p_g = 0.10$. Convergence was set as the point when the best model was the same for 50 generations.

For each form of excretion, in each animal class, the 6 models selected from different variable pools by the genetic algorithm were compared and the one with the lowest BIC was chosen. The BIC determination was based on maximum likelihood fits and all models were then refit using restricted maximum-likelihood to report unbiased estimates of the variance components. Model intercepts are reported in terms of the uncentered independent variables for which the standard errors were derived through the delta-method (DeGroot and Schervish, 2010).

Model Evaluation and Comparison

The final predictive power of the selected models was assessed using a *K*-fold cross-validation (Picard and Cook, 1984; Efron and Tibshirani, 1994). In this method, the data are iteratively divided into training and testing data sets to obtain estimates of the prediction error from data that was not used in model development. The data was folded by study, resulting in 59 folds divided by the appropriate animal class, and the prediction error was estimated by the root mean square prediction error (**RMSPE**). Thus, the cross-validation technique used is similar to the "leave-one-out" technique in which a whole study was left out of the training set instead of a single observation. Here the RMSPE was calculated according to the following formula (Bibby and Toutenburg, 1977):

$$\text{RMSPE} = \left[\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \right)^{\mathrm{T}} \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \right) n^{-1} \right]^{1/2}$$

where $\mathbf{y}$ is the vector of observations; $\mathbf{X}$ is the matrix of explanatory variables; $\hat{\boldsymbol{\beta}}$ is the vector of fixed-effect parameter estimates; n is the number of observations in each fold; and the superscript T indicates the transpose function. The final prediction error was calculated as a weighted average in which each RMSPE was weighted by the number of observations in the fold and is presented both in the units of the predicted value and as a percentage of the mean observed value. In its original units, the RMSPE is an indicator of the average uncertainty of prediction and when expressed as a percentage allows for comparison of prediction accuracy across models.

To assess the models' application in predicting observations of more recent years, the models were refit excluding the last 6 yr of observations and then evaluated using the RMSPE with the independent data subset of observations from the last 6 yr. Descriptions of this data subset are given in Table 2. In addition, to compare model performance with extant models, the RMSPE of 3 FN models (Castillo et al., 2000; Huhtanen et al., 2008; Kebreab et al., 2010), 3 UN models (Castillo et al., 2000; Huhtanen et al., 2008; Kebreab et al., 2010), and 2 TN models (Nennich et al., 2005; Yan et al., 2006) was calculated with both the complete data and the most recent data subset. To assess the models' potential for practical application, the TN prediction equations used in the national GHG inventory (EPA, 2014) were evaluated for prediction accuracy with both the complete and most recent data. In addition to prediction error of the extant models, prediction bias was assessed through decomposition of the prediction error into mean bias (error of central tendency), systematic bias (error due to regression), and the random error (dispersion error) as a percent of the MSPE (Bibby and Toutenburg, 1977).

RESULTS AND DISCUSSION

Methodology

Assessment of the amount and form of N escaping ruminant production facilities requires accurate prediction equations. Most extant models were developed from a small pool of potential covariates and selected based on either biological knowledge or a stepwise selection method; both of which are subject to criticism. A more thorough search of available covariates and model parameter space is likely to improve prediction ability. However, when the number of possible covariates is large, model selection can be a time- and computation-intensive process. The genetic algorithm used in our study facilitated the selection process and reduced the time required to identify the best possible model based on a given criteria. In the present application, the algorithm converged to the models with the lowest BIC in less than 100 iterations.

Model Development

The models developed in our study through the selection process described previously are presented in Table 3. Two tiers of models were selected for each class of animals to accommodate variable input requirements. Nitrogen intake was selected as the best predictor compared with DMI for all 3 animal classes in tier 1 models for UN and TN. In contrast, DMI was selected as the best predictor for FN and MN in lactating cows, whereas NI was selected for FN prediction in heifers and dry cows and steers. The selection of DMI as a better predictor of FN instead of NI in lactating cows was not surprising. Although greater NI and DMI increase rumen microbial growth and N absorption across the digestive tract, greater DMI will also increase the rate of digesta passage and secretion of digestive enzymes resulting in greater fecal DM and FN (Swanson, 1977; Dijkstra et al., 2013). As expected, all measures of excretion, for all animal classes and tiers had a positive linear relationship with the intake parameter (NI or DMI).

A measurement of the energy content of the diet was included in all selected tier 2 models except equation 16 in Table 3, predicting UN excretion in steers. The estimated regression coefficients reveal a consistent positive relationship of energy content with UN (and MN for lactating cows), and a consistent negative relationship with FN and TN. The apparent increase in UN with increasing diet energy content is in contrast to findings by Kebreab et al. (2010). However, the univariate models reported by Kebreab et al. (2010) used ME intake (kJ/d) or metabolizability (kJ of ME/kJ of gross energy) covariates instead of ME or DE content (kJ/kg of DM) as reported in our study. In the present data set, ME intake was highly correlated (r = 0.92) to NI, and thus would be expected to increase all forms of N excretion as shown by Kebreab et al. (2010). Metabolizability, conversely, was highly correlated to ME and DE content (kJ/kg of DM) and the negative relationship of UN and metabolizability reported by Kebreab et al. (2010) was not consistent with the present findings. However, Kebreab et al. (2010) reported the standard error of the parameter estimate for predicting UN from metabolizability was 46% of the mean estimate. This is much larger than current standard errors as a proportion of their mean estimates, which was perhaps a

Table 2. Units, means, SD, and range of data collected from 1990 to 1995 used to evaluate the selected models

Variable[1]	Unit	Lactating cows (n = 82)				Heifers and dry cows (n = 52)				Steers (n = 30)			
		Mean	SD	Minimum	Maximum	Mean	SD	Minimum	Maximum	Mean	SD	Minimum	Maximum
UN	g/d	232	61.5	123	350	122	24.3	74.0	171	65.3	23.1	36.0	115
FN	g/d	223	56.2	105	334	48.5	8.58	35.7	72.4	40.3	8.36	26.3	57.5
TN	g/d	455	109	230	641	170	27.1	131	227	106	28.6	62.8	169
MN	g/d	161	48.0	47.9	267	—	—	—	—	—	—	—	—
Milk	kg/d	35.0	9.55	6.47	56.6	—	—	—	—	—	—	—	—
NI	g/d	632	161	294	932	169	22.7	124	212	127	29.6	85.6	197
DMI	kg/d	21.0	4.39	10.3	28.2	6.01	0.71	4.64	7.40	5.72	0.89	4.42	7.43
DE	MJ/kg of DM	12.9	0.83	11.1	14.9	14.0	0.94	10.7	15.5	13.5	0.72	12.3	15.0
ME	MJ/kg of DM	11.1	0.81	9.57	13.1	11.5	0.81	8.84	13.1	11.8	0.67	10.7	13.2
DM	%	48.6	5.81	34.2	58.2	51.2	3.78	38.5	56.6	88.3	0.88	86.5	89.7
EE	%	3.90	0.89	2.18	6.52	4.31	1.49	2.07	7.64	2.97	0.83	1.64	4.04
NDF	% of DM	37.0	8.66	25.6	59.3	35.5	5.57	28.0	50.1	24.8	3.48	19.9	32.8
ADF	% of DM	23.4	3.51	16.2	34.5	22.2	3.97	17.1	36.3	11.2	1.36	7.50	15.2
Lignin	% of DM	6.40	0.96	4.30	9.01	5.34	1.48	3.15	8.76	3.35	0.74	2.06	4.96
Ash	% of DM	7.48	0.51	6.39	8.65	8.44	1.98	6.20	13.3	5.14	0.49	3.92	6.15
CP	% of DM	18.7	1.46	16.4	21.2	17.6	1.01	14.5	19.2	13.9	2.27	10.8	17.1
Forage	% of DM	71.1	24.6	50.0	100	67.8	22.8	50.0	100	0.25	0	0.25	0.25
DIM	d	142	70	41	299	—	—	—	—	—	—	—	—
BW	kg	578	102	355	759	625	138	339	797	396	33.5	331	457

[1]UN = urinary nitrogen, FN = fecal nitrogen, TN = total manure nitrogen, MN = milk nitrogen, NI = N intake, DE = digestible energy, and EE = ether extract.

Table 3. Model selection results and variance estimates[1]

Tier	Equation no.	Estimate[2]	σ_R^2	σ_a^2	σ_s^2
Lactating cattle					
1	1	$UN = 12.0(5.80) + 0.333(0.0106) \times NI$	610	307	420
2	2	$UN = -260(2.81) + 0.293(0.0101) \times NI + 3.99(1.11) \times DE + 0.906(0.158) \times NDF + 8.39(0.578) \times CP + 0.161(0.0108) \times DIM + 0.370(0.821) \times MBW$	411	101	225
1	3	$FN = -18.5(3.59) + 10.1(0.169) \times DMI$	178	39.6	195
2	4	$FN = 72.7(2.34) - 11.8(0.609) \times ME - 0.437(0.0979) \times NDF + 3.52(0.258) \times CP + 0.161(0.478) \times ForR + 9.32(0.148)\,DMI - 0.0184(0.00554) \times DIM$	117	24.6	195
1	5	$TN = 20.3(4.72) + 0.654(0.00926) \times NI$	567	126	212
2	6	$TN = 2.16(2.88) + 0.631(0.0105) \times NI - 12.8(1.07) \times ME + 0.392(0.160) \times NDF + 4.81(0.540) \times CP + 0.137(0.0104) \times DIM + 0.426(0.0764) \times MBW$	400	67.7	276
1	7	$MN = -19.0 + 8.13(0.245) \times DMI$	370	103	286
2	8	$MN = 5.62(1.65) + 4.62(0.714) \times ME - 0.389(0.104) \times NDF + 1.70(0.318) \times CP + 6.66(0.198) \times DMI - 0.194(0.00714) \times DIM - 0.275(0.0634) \times MBW$	163	113	70.6
Heifers and nonlactating cows					
1	9	$UN = 14.3(3.18) + 0.510(0.0121) \times NI$	168	66.0	215
2	10	$UN = -43.8(2.94) + 0.429(0.0133) \times NI - 2.10(0.691) \times ME + 0.292(0.113) \times ADF + 3.35(0.313) \times CP + 0.284(0.0400) \times MBW$	166	13.8	160
1	11	$FN = 0.345(1.73) + 0.317(0.00638) \times NI$	31.3	6.73	40.8
2	12	$FN = 107(1.34) + 0.359(0.00567) \times NI - 4.18(0.303) \times ME - 0.141(0.0283) \times DM - 0.246(0.0542) \times ADF - 1.00(0.166) \times ASH - 2.32(0.137) \times CP - 0.0792(0.0177) \times MBW$	26.6	5.77	36.7
1	13	$TN = 15.1(2.50) + 0.828(0.0106) \times NI$	145	13.6	111
2	14	$TN = 49.5(2.57) + 0.793(0.0115) \times NI - 6.04(0.0501) \times ME + 0.825(0.269) \times CP + 0.190(0.0332) \times MBW$	127	3.06	129
Steers					
1	15	$UN = 6.80(3.78) + 0.405(0.0166) \times NI$	194	1.89	77.2
2	16	$UN = -71.2(3.32) + 0.265(0.0181) \times NI + 3.76(0.344) \times CP + 0.468(0.0522) \times MBW$	143	5.60	23.4
1	17	$FN = 0.506(2.30) + 0.352(0.00902) \times NI$	57.3	0	32.4
2	18	$FN = 109.6(1.99) + 0.327(0.00724) \times NI - 6.41(0.334) \times ME - 0.276(0.531) \times DM - 1.37(0.243) \times LIG - 1.82(0.274) \times ASH + 0.0398(0.0230) \times MBW$	26.4	2.90	25.2
1	19	$TN = 6.91(2.60) + 0.759(0.0126) \times NI$	116	0	31.0
2	20	$TN = 11.5(2.49) + 0.650(0.0152) \times NI - 4.47(0.612) \times ME + 1.77(0.278) \times CP + 0.432(0.0423) \times MBW$	88.0	18.1	3.1

[1]The variance components of the mixed model include the residual variance $\left(\sigma_R^2\right)$, the animal variance $\left(\sigma_a^2\right)$, and the study variance $\left(\sigma_s^2\right)$.

[2]Parameter estimates with SE in parentheses. UN = urine nitrogen, FN = fecal nitrogen, MN = milk nitrogen, TN = total manure nitrogen, NI = nitrogen intake (g/d), DE = digestible energy content (kJ/kg DM), MBW = metabolic body weight (body weight$^{3/4}$, kg), LIG = lignin content (% DM), ForR = proportion of forage in the diet, ASH = ash content (% DM).

Table 4. Evaluation of selected models using a *K*-fold cross-validation (USDA cross) and splitting the data to be evaluated with the data from the most recent 6 yr (USDA 1990–1995)

Equation no.	Predicted variable[1]	USDA cross[2]		USDA 1990–1995[2]	
		RMSPE (g/d)	RMSPE (% of mean N excreted)	RMSPE (g/d)	RMSPE (% of mean N excreted)
Lactating cows					
1	UN	35.1	23.7	29.5	12.7
2	UN	26.6	18.0	24.6	10.6
3	FN	18.0	12.6	40.6	18.2
4	FN	16.4	11.4	37.1	16.6
5	TN	28.7	9.80	37.0	8.14
6	TN	25.9	8.87	36.4	8.00
7	MN	24.6	21.9	24.1	14.9
8	MN	17.5	15.6	19.6	12.2
Heifers and dry cows					
9	UN	20.6	22.2	28.3	23.2
10	UN	17.1	18.4	23.7	19.4
11	FN	8.4	16.2	6.34	13.1
12	FN	7.1	14.2	5.93	12.2
13	TN	16.2	11.4	22.4	13.2
14	TN	14.4	10.1	23.9	14.1
Steers					
15	UN	15.5	26.5	15.2	23.2
16	UN	12.7	21.7	12.9	19.8
17	FN	8.2	17.6	8.59	22.0
18	FN	6.6	14.2	6.10	15.1
19	TN	11.6	11.0	9.38	8.89
20	TN	9.7	9.2	6.73	6.37

[1]UN = urine nitrogen, FN = fecal nitrogen, MN = milk nitrogen, TN = total manure nitrogen.
[2]The square root of the mean square prediction error (RMSPE) expressed both in grams of N excretion per day and as a percentage of the mean observed excretion.

symptom of the smaller data set used by Kebreab et al. (2010). A simple plot of UN excretion versus energy content (not shown) confirms the proposed positive parameter estimate indicating an increase in UN with energy content.

For lactating cows, NDF, CP, and DIM were included in the selected tier 2 models for all forms of excretion and MN secretion. Neutral detergent fiber had a positive linear relationship with UN and TN and a negative relationship with FN and MN. As with intake, diet CP had a positive relationship with all dependent variables. Metabolic BW was selected for all tier 2 models predicting UN and TN but not for all FN equations, suggesting that BW has a stronger relationship with UN than FN. For all excretion and animal types, a reduction in residual and total variance was noted when moving from tier 1 to tier 2 models. This reduction in variance was also reflected in the cross-validation RMSPE reported in Table 4, showing that, overall, prediction accuracy increased when more dietary information was included.

Model Evaluation

Prediction Models. A cross-validation procedure was used to evaluate the overall predictive ability of the fitted models, as described previously. Furthermore, the data was divided into 2 subsets containing the most recent 6 yr in the database and those collected previous to the last 6 yr to investigate whether the models based on earlier years were still relevant in the most recent 6 yr given the 3-decade collection period. As expected, UN prediction equations resulted in the greatest prediction errors when evaluated with both the cross-validation RMSPE and the RMSPE of the data from the most recent 6 yr (Table 4). Total manure N consistently resulted in the lowest prediction errors among the 3 forms of excretion for all animal classes. Fecal N prediction error was intermediate compared with UN and TN. Differences in the RMSPE between the 2 evaluation techniques (cross-validation and splitting the data by time period) were variable across animal and excretion types, ranging ±5% of the mean. The distribution of the variables of interest in the subset of data collected after 1990 is different from the complete data and more closely represents recent trends in animal size and diets (Table 2). As a result, the nonuniform change in the direction of prediction error when evaluated with this subset suggests prediction accuracy is not dependent on year, and the proposed equations are applicable to present practices given the required covariates are jointly within the bounds of the data used for model development (Table 1).

Comparison with Extant Models. The results of the evaluation of extant models are presented in Table 5. All extant UN and FN models evaluated in our study for lactating cows resulted in larger prediction errors than the proposed models when evaluated with the complete data and most recent data. This suggests that if separate estimation of FN and UN is desired, use of the proposed models is recommended. The 2 extant prediction equations for TN (equations 27 and 28) listed in Table 5 performed better when evaluated with the most recent data than with the complete data. Under the complete data evaluation, the model proposed by Yan et al. (2006; equation 28) was within 1 g of the prediction error of the proposed models (equations 5 and 6), but the model by Nennich et al. (2005; equation 27) was approximately 20 g larger than both equation 28 and the proposed prediction equations for TN excretion from lactating cows (equations 5 and 6). Under evaluation with the most recent data, both equations 27 and 28 performed slightly better than the proposed models, with a prediction error approximately 2 g smaller than that of equations 5 and 6, suggesting they were all equally suited for predicting TN from lactating cows. The data used for the analysis was greater and statistical procedure for model selection was different than most extant models, and this might have contributed to lower uncertainty associated with models developed in our study.

All extant models showed either significant mean (error of central tendency) or systematic (dispersion error) bias, with the exception of equations 28 and 29, when they were evaluated with the complete data set (Table 5). There was an increase in mean bias when the models were evaluated with data from the most recent 6 yr for all but one model (equation 27). In particular, models predicting UN indicate large mean bias when evaluated with the complete and most recent data sets (equations 24–26; Table 5). The TN model by Nennich et al. (2005; equation 27) had large mean bias when evaluated with the complete data set; however, this decreased dramatically when evaluated with the most recent 6 yr of data. The decompositions of error for the non-EPA, TN models (equations 27–28) were overall better than the error decompositions of the other forms of excretion.

Practical Application

The N source for agricultural N_2O emissions is predominantly livestock manure N worldwide (Gerber et al., 2013), and the combination of synthetic and manure fertilizer in the United States (EPA, 2014). Nitrous oxide is generated by microorganisms as an intermediate in both nitrification and denitrification processes in aerobic and anaerobic environments while manure is stored, treated, transported and applied to soil. Nitrogenous emissions, including ammonia volatilization, can occur directly during manure handling and storage. Indirect emissions occur from N that originally escaped through volatilization, leaching, or run-off or was applied for soil amendment (Owen et al., 2014). In addition to N_2O escape to the atmosphere, manure N can result in eutrophication of surface waters, acid rain, and respiratory damage (McCubbin et al., 2002; Diaz and Rosenberg, 2008). Reducing the intensity of N excretion (i.e., amount of manure N produced per kilogram of milk or meat) is an important way to mitigate the environmental effects of animal agriculture (Montes et al., 2013).

Specific estimation of FN and UN is useful for prediction of the environmental fate of N that escapes cattle-production facilities. For example, urea N in urine is volatilized to ammonia when it comes into contact with urease in feces and subsequent deposition of this ammonia can result in eutrophication of water sources or indirect N_2O emissions. Most emissions inventories, however, are based on estimation of TN. The models used by the EPA (2014) for predicting TN from lactating cows, heifers and dry cows, and steers are presented in Table 5 (equations 20, 30, and 31, respectively) along with their evaluation results. These models resulted in large prediction errors both when evaluated with the complete and restricted data; with the exception of the equation for heifers and dry cows, which performed fairly well when evaluated with the most recent data. The general strategy of the EPA methodology is to estimate TN as the difference in NI and either the N deposited in tissue or secreted in milk. In lactating cows this reduces to TN = NI – MN by assuming retained or mobilized N is zero. Over the course of a lactation this, in general, is an appropriate assumption; however, on a daily basis, this assumption does not hold, as the retained N during lactation can be quite variable (Andrew and Erdman, 1995; Komaragiri and Erdman, 1997). Thus, application of this model should be limited to estimates of annual herd TN excretion and additional estimates for which the zero N retention assumption is reasonable.

The models can also be used in developing quantification protocols for producers to assess the effect of dietary manipulations on N partitioning to products and excreta. However, as with all regression models, caution should be taken in their application for prediction. Inference about the effect of dietary and animal parameters on N excretion can be drawn from the proposed models; however, the regression parameters must be interpreted from within the context of the model and development data. Each parameter estimate

Table 5. Evaluation of extant fecal N (FN), urinary N (UN), and total N (TN) excretion prediction equations of lactating cows and their evaluation with the complete data (USDA complete) and the data from the most recent 6 yr (USDA 1990–1995) using the mean square prediction error (MSPE), its square root (RMSPE), and its decomposition into the error of central tendency (ECT), error of regression (ER), and dispersion error (ED)[1]

Equation no.	Model	Citation	USDA complete					USDA 1990–1995				
			RMSPE (g/d)	RMSPE (% of mean N excretion)	ECT	ER	ED	RMSPE (g/d)	RMSPE (% of mean N excretion)	ECT	ER	ED
					(% of MSPE)					(% of MSPE)		
21	FN = 52.3 + 0.21 × NI	Castillo et al. (2000)	31.8	21.1	5.49	32.2	62.3	69.5	27.2	70.0	0.37	29.6
22	FN = −21 + 6.73 × DMI + 0.101 × NI	Huhtanen et al. (2008)	28.3	18.9	28.5	6.19	65.3	62.3	24.4	69.1	0.15	30.7
23	FN = 244 + 0.25 × NI − 346 × q	Kebreab et al. (2010)	23.8	15.8	2.22	26.6	71.2	46.2	18.1	55.9	4.89	39.2
24	UN = 30.4 × $e^{0.0036 \times NI}$	Castillo et al. (2000)	73.7	48.4	4.00	62.8	33.8	190	79.3	62.0	33.7	4.25
25	UN = 40 + 0.879 × NI − 9 × DMI − 3.9 × MILK	Huhtanen et al. (2008)	41.3	27.1	43.6	7.05	49.4	62.3	26.0	63.4	9.34	26.9
26	UN = 47.8 + 0.56 × NI − 71.4 × ME	Kebreab et al. (2010)	44.7	29.3	46.7	0.35	53.0	59.6	24.9	60.3	1.44	38.31
27	TN = 0.196 × BW + 84.1 × CP × DMI	Nennich et al. (2005)	55.8	18.4	52.5	11.7	35.8	34.3	6.9	11.1	1.23	87.7
28	TN = −18 + 0.742 × NI + 0.072 × BW − 1.486 × MILK	Yan et al. (2006)	28.4	9.39	6.62	4.00	89.4	35.9	7.3	14.8	7.90	77.3
29	$TN = \dfrac{GE}{18.45} \times \dfrac{CP/100}{6.25} - \left[\dfrac{WG \times \left[268 - \dfrac{(7.03 \times NE_g)}{WG}\right]}{1{,}000} \times \dfrac{1}{6.25} + \dfrac{0.035 \times MILK}{6.38}\right]$	EPA (2014)	69.1	22.8	3.66	0.00	96.3	70.0	15.0	52.5	0.53	47.0
30	$TN = \dfrac{GE}{18.45} \times \dfrac{CP/100}{6.25} - \left[\dfrac{WG \times \left[268 - \dfrac{(7.03 \times NE_g)}{WG}\right]}{1{,}000} \times \dfrac{1}{6.25}\right]$	EPA (2014)	39.2	27.4	18.3	10.6	71.1	32.4	30.7	85.7	0.00	14.3
31	$TN = \dfrac{GE}{18.45} \times \dfrac{CP/100}{6.25} - \left[\dfrac{WG \times \left[268 - \dfrac{(7.03 \times NE_g)}{WG}\right]}{1{,}000} \times \dfrac{1}{6.25}\right]$	EPA (2014)	33.6	32.0	5.55	23.2	71.2	46.7	44.3	38.9	8.81	52.3

[1]Evaluation of the US Environmental Protection Agency model for predicting TN excretion for lactating cows (equation 29), heifers and dry cows (equation 30), and steers (equation 31) is also given.

indicates the average change in the dependent variable for a unit change of the accompanying covariate, if all other covariates are held constant.

CONCLUSIONS

National and international inventories of GHG emissions from livestock require accurate N excretion prediction equations available to estimate direct and indirect N_2O emissions. Aggregate estimates of N_2O emissions unavoidably compound errors associated with each level of estimation. Thus, models with the lowest prediction errors should be applied when available. The proposed equations for manure N excretion from lactating cows performed better or similar to extant models; however, no comparable equations for steers or heifers and dry cows were found. In addition, the equations presented for UN and FN had substantially smaller prediction errors than extant models. Separate estimation of UN and FN provides a more detailed description of N excretion. The estimates can be used to further improve assessment of the environmental effects of N excretions from cattle in connection with process-based or empirical models predicting emissions from manure storage and field application.

ACKNOWLEDGMENTS

REFERENCES

Andrew, S. M., and A. Erdman. 1995. Prediction of body composition of dairy cows at three physiological stages from deuterium oxide and urea dilution. J. Dairy Sci. 78:1083–1095.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2013. lme4: Linear mixed-effects models using Eigen and S4. Vol. R package version 1.0–5. http://cran.r-project.org/web/packages/lme4/index.html.

Belsley, D. A., E. Kuh, and R. E. Welsch. 2004. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley and Sons, Hoboken, NJ.

Bibby, J., and H. Toutenburg. 1977. Prediction and Improved Estimation in Linear Models. John Wiley and Sons, London, UK.

Castillo, A. R., E. Kebreab, D. E. Beever, and J. France. 2000. A review of efficiency of nitrogen utilisation in lactating dairy cows and its relationship with environmental pollution. J. Anim. Feed Sci. 9:1–32.

DeGroot, M. H., and M. J. Schervish. 2010. Probability and Statistics. 4th ed. Addison Wesley, Boston, MA.

Diaz, R. J., and R. Rosenberg. 2008. Spreading dead zones and consequences for marine ecosystems. Science 321:926–929.

Dijkstra, J., O. Oenema, J. W. van Groenigen, J. W. Spek, A. M. van Vuuren, and A. Bannink. 2013. Diet effects on urine composition of cattle and N_2O emissions. Animal 7(Suppl. 2):292–302.

Efron, B., and R. J. Tibshirani. 1994. An Introduction to the Bootstrap. Taylor & Francis, New York, NY.

EPA. 2014. Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2012. United State Environmental Protection Agency (EPA), Washington, DC. Accessed Jul. 1, 2014. http://www.epa.gov/climatechange/emissions/usinventoryreport.html.

Forrest, S. 1993. Genetic algorithms: Principles of natural selection applied to computation. Science 261:872–878.

Gerber, P. J., H. Steinfeld, B. Henderson, A. Mottet, C. Opio, J. Dijkman, A. Falcucci, and G. Tempio. 2013. Tackling Climate Change Through Livestock—A Global Assessment of Emissions and Mitigation Opportunities. Food and Agriculture Organization of the United Nations, Rome, Italy.

Hendrickx, J. 2012. perturb: Tools for Evaluating Collinearity. Vol. R package version 2.05. http://cran.r-project.org/web/packages/perturb/index.html.

Higgs, R. J., L. E. Chase, and M. E. Van Amburgh. 2012. Development and evaluation of equations in the Cornell Net Carbohydrate and Protein System to predict nitrogen excretion in lactating dairy cows. J. Dairy Sci. 95:2004–2014.

Huhtanen, P., J. I. Nousiainen, M. Rinne, K. Kytölä, and H. Khalili. 2008. Utilization and partition of dietary nitrogen in dairy cows fed grass silage-based diets. J. Dairy Sci. 91:3589–3599.

IPCC. 2006. 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Intergovernmental Panel on Climage Change: Insitute for Global Environmental Strategies. Hayama, Japan.

Kebreab, E., A. B. Strathe, J. Djikstra, J. A. N. Mills, C. K. Reynolds, L. A. Crompton, T. Yan, and J. France. 2010. Energy and protein interactions and their effect on nitrogen excretion in dairy cows. Pages 417–425 in Proc. 3rd EAAP International Symposium on Energy and Protein Metabolism. M. Crovetto, ed. Wageningen Publishers, Wageningen, The Netherlands.

Komaragiri, M. V. S., and R. A. Erdman. 1997. Factors affecting body tissue mobilization in early lactation dairy cows. 1. Effect of dietary protein on mobilization on body fat and protein. J. Dairy Sci. 80:929–937.

Li, C., W. Salas, R. Zhang, C. Krauter, A. Rotz, and F. Mitloehner. 2012. Manure-DNDC: A biogeochemical process model for quantifying greenhouse gas and ammonia emissions from livestock manure systems. Nutr. Cycl. Agroecosystems 93:163–200.

McCubbin, D. R., B. J. Apelberg, S. Roe, and F. Divita. 2002. Livestock ammonia management and particulate-related health benefits. Environ. Sci. Technol. 36:1141–1146.

Montes, F., R. Meinen, C. Dell, A. Rotz, A. N. Hristov, J. Oh, G. Waghorn, P. J. Gerber, B. Henderson, H. P. S. Makkar, and J. Dijkstra. 2013. Mitigation of methane and nitrous oxide emissions from animal operations: II. A review of manure management mitigation options. J. Anim. Sci. 91:5070–5094.

Nennich, T. D., J. H. Harrison, L. M. VanWieringen, D. Meyer, A. J. Heinrichs, W. P. Weiss, N. R. St-Pierre, R. L. Kincaid, D. L. Davidson, and E. Block. 2005. Prediction of manure and nutrient excretion from dairy cattle. J. Dairy Sci. 88:3721–3733.

NRC. 2001. Nutrient Requirements of Dairy Cattle. 7th rev. ed. Natl. Acad. Sci., Washington, DC.

Owen, J. J., E. Kebreab, and W. Silver. 2014. Review of Emissions and Mitigation Potential of Animal Manure Management and Land Application of Manure. Nicholas Institute for Environmental Policy Solutions. Duke University, Durham, NC.

Picard, R. R., and R. D. Cook. 1984. Cross-validation of regression models. J. Am. Stat. Assoc. 79:575–583.

Swanson, E. W. 1977. Factors for computing requirements of protein for maintenance of cattle. J. Dairy Sci. 60:1583–1593.

Tamminga, S. 1996. A review on environmental impacts of nutritional strategies in ruminants. J. Anim. Sci. 74:3112–3124.

UN. 1992. United Nations Framework Convention on Climate Change. United Nations (UN), New York, NY.

Ward, E. J. 2008. A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. Ecol. Modell. 211:1–10.

Wilkerson, V. A., D. R. Mertens, and D. P. Casper. 1997. Prediction of excretion of manure and nitrogen by Holstein dairy cattle. J. Dairy Sci. 80:3193–3204.

Yan, T., J. P. Frost, R. E. Agnew, R. C. Binnie, and C. S. Mayne. 2006. Relationships among manure nitrogen output and dietary and animal factors in lactating dairy cows. J. Dairy Sci. 89:3981–3991.