

# Multivariate modeling for retained protein and lipid<sup>1</sup>

Luis Eduardo Moraes<sup>2,\*</sup>

Department of Animal Sciences, The Ohio State University, Columbus, OH 43210

**ABSTRACT:** Energy efficiencies and maintenance parameters have been traditionally estimated with a linear regression model that treated metabolizable energy intake as the dependent variable and protein and lipid depositions as the independent variables. Several studies have described the statistical issues associated with this approach, such as the reverse role of dependent and independent variables and a potential multicollinearity issue due to the high correlation between protein and lipid depositions. Biased regression techniques have been proposed to minimize the harmful effects of multicollinearity on the estimates of energy efficiencies. These approaches, however, only partially addressed the issues described for the linear regression approach. A first multivariate approach was developed by L. J. Koong in the 1970s, who estimated the energy parameters using a set of simultaneous equations. This multivariate approach has been considerably

extended in the past two decades with the complete characterization of model's biological interpretation under different feeding conditions, the simultaneous estimation of maintenance requirements, the extension of the model to a mixed-effects framework, and the implementation of a Bayesian framework for model fitting. The multivariate approach has been successfully applied to model energy deposition and partitioning by mice, pigs, salmon, and rainbow trout. However, multivariate models are, in general, harder to fit than linear regression models due to 1) larger number of parameters, 2) issues with parameter identifiability, and 3) overall lack of algorithm convergence. Therefore, with the recent availability of easy to use and efficient computer packages for model fitting, the use of a Bayesian framework seems to be an attractive approach for fitting multivariate models describing protein and lipid deposition.

**Key words:** efficiency, growth, lipid, protein, requirement, simultaneous equations

© The Author(s) 2019. Published by Oxford University Press on behalf of the American Society of Animal Science.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Transl. Anim. Sci. 2019.3:1040–1047  
doi: 10.1093/tas/txz017

## INTRODUCTION

Feeds often represent the largest fraction of production costs in most livestock systems. Minimizing feeding costs for profitable animal production requires a comprehensive characterization

of nutrient availability in feeds and a precise determination of animal's nutrient requirements for maintenance and production functions. Current feeding systems for cattle represent animals' energy requirements and the energy availability in feeds on a net energy basis (Lofgreen and Garrett, 1968; Moe et al., 1972). The connection between net energy availability in feeds and the animal's requirements relies on quantification of energy efficiencies and determination of the maintenance requirements. The efficiency with which metabolizable energy (ME) from the diet is utilized for maintenance and

<sup>1</sup>Research was supported by state and federal funds appropriated to the Ohio Agricultural Research and Development Center, The Ohio State University.

<sup>2</sup>Corresponding author: [ferrazdiasdemoraes.1@osu.edu](mailto:ferrazdiasdemoraes.1@osu.edu)

Received October 2, 2018.

Accepted March 25, 2019.

growth has stimulated scientific research for decades. Likewise, the quantification of maintenance requirements for net energy ( $NE_m$ ) and metabolizable energy ( $ME_m$ ) has been extensively explored by research groups worldwide (e.g., [Blaxter and Wainman, 1966](#); [Moraes et al., 2014](#)). Traditionally, the estimation of these energy efficiencies has relied on the use of a multiple regression model:

$$MEI/MBS = \beta_0 + \beta_1(LG/MBS) + \beta_2(PG/MBS)$$

where MEI is the ME intake (Mcal or MJ), MBS is the metabolic body size ( $kg^{0.75}$ ), LG is the energy gain as lipid (Mcal or MJ), and PG is the energy gain as protein (Mcal or MJ). In this approach, energy parameters, such as maintenance and efficiencies, are estimated as function of the parameters. For example, the efficiency of utilizing ME for protein gain is computed as  $k_p = 1/\beta_2$ . Over the years, some variant of equation (1) has been the basis for estimation of maintenance and efficiency parameters in numerous publications, despite the comprehensive critique of this approach by [Koong \(1977\)](#). For instance, the reverse definition of the dependent vs. independent variables and issues with multicollinearity in equation (1) were described in detail at least three decades ago ([Koong, 1977](#); [Bernier et al., 1987](#)), but this approach is still in use worldwide. It is important to note that although there are issues described by these authors underlying the general structure of this model, this univariate approach provided a method for estimating energy parameters at a time when computing power was probably a limiting factor for estimating more complex models.

Equation (1) can be seen as a univariate approach, in the sense that it has one dependent variable related to a set of independent variables. A series of multivariate models, with protein deposition (PD) and lipid deposition (LD) as dependent variables, has been proposed to overcome the limitation of equation (1). In this context, the objective of this article is to describe this univariate approach and to explore the use of multivariate models as alternatives for estimating energy efficiencies, maintenance requirements, and modeling LD and PD. Our secondary objective is to provide a historical timeline of major developments in techniques to model protein and lipid deposition in growing animals while describing both advantages and disadvantages of each approach.

### Univariate Approach

Generally, equation (1) can be described as a linear regression model with multiple independent variables:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad (2)$$

where  $y_i$  is the dependent variable for the  $i$ th observation ( $i = 1, \dots, N$ ),  $\beta_0$  is the intercept,  $\beta_1, \dots, \beta_p$  are the regression coefficients describing the relationships between dependent and independent variables  $x_{1i}, \dots, x_{pi}$  and  $\varepsilon_i$  is the error, often assumed to be independent and identically distributed as  $\varepsilon_i \sim N(0, \sigma^2)$ . Parameters in this model can be estimated by ordinary least squares although, under the assumptions just described, maximum likelihood estimators would produce the same solutions ([Kutner et al., 2004](#)). The estimates for equation (2) are obtained through the following minimization problem:

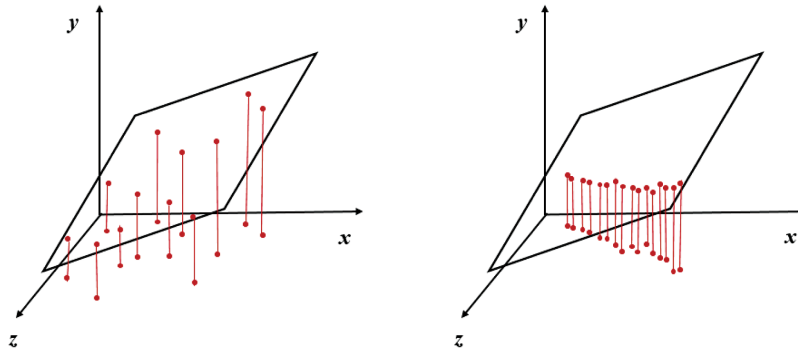
$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\} \quad (3)$$

where  $\operatorname{argmin}$  is the argument of the minimum,  $\hat{\beta}$  is the vector of parameter estimates, often referred in the statistics literature as least square estimates (LSE). The LSE have many desirable properties, such as when the errors are uncorrelated, have zero expectation and equal variances, the estimates in equation (2) are unbiased and have minimum variance among all unbiased linear estimators ([Kutner et al., 2004](#)).

The LSE, however, can have large sampling variability if independent variables are highly correlated, a problem known as multicollinearity. With high correlation among independent variables, parameter estimates may vary dramatically from one sample to another, and the interpretation of regression coefficients as the marginal changes may be not valid ([Kutner et al., 2004](#)). A geometric representation of the effects of multicollinearity was comprehensively described by [Slinker and Glantz \(1985\)](#). For example, consider the case where the dependent variable  $y_i$  is to be modeled through the use of independent variables  $x_i$  and  $z_i$ :

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i \quad (4)$$

It is easy to see that equation (4) is a special case of equation (2) such that  $x_{1i} = x_i$  and  $x_{2i} = z_i$ . To obtain the LSE, we need to identify the plane that minimizes the squared distances with the points in space ([Slinker and Glantz, 1985](#)). This process is presented in [Figure 1](#). The plot on the left describes a situation with low degree of correlation among independent variables. Observations provide a good “base” for the plane and for the estimation of stable model parameters ([Slinker and Glantz, 1985](#)). Contrariwise, the plot on the right describes the situation with high collinearity between  $x$  and  $z$ .



**Figure 1.** Visual representation of multicollinearity in a linear regression model. The plot on the left represents the scenario where  $x$  and  $z$  have low correlation, whereas the plot on the right represents the case where  $x$  and  $z$  are highly collinear. Figure adapted from [Slinker and Glantz \(1985\)](#). The plot on the left describes the case when  $x$  and  $z$  provide a good “base” for the plane because they are distributed over the entire region. In this scenario, observations provide a stable “base” for the estimation of regression parameters. The plot on the right describes the situation with highly collinear  $x$  and  $z$ . The base formed by observations is too narrow to uniquely estimate the plane position, resulting in unstable estimates for the regression parameters. The narrow base is described by [Slinker and Glantz \(1985\)](#) as a “base” that allows the plane to “wobble around” and being unstable.

Observations, in this case, provide a base that is excessively narrow for the estimation of the plane location and, consequently, for the estimation of stable parameters. Thus, parameter estimates will be unstable and suffer from high sampling variability. The situation reaches the extreme when  $x$  and  $z$  are perfectly collinear and  $y$  can be described by a line in the plane that can be determined equally well by more than one set of positions ([Slinker and Glantz, 1985](#)). In this extreme case, model parameters may not be estimated simultaneously.

From an energetics standpoint, multicollinearity between the two independent variables in equation (1) has the potential to prevent stable and accurate estimates of energy efficiencies and maintenance parameters. That is, if protein and lipid depositions are highly correlated, the base for the geometric plane is probably narrow, and the estimates of energy efficiencies and maintenance will probably suffer from large sampling variability and have poor interpretability.

**Ridge regression approach.** A variety of techniques exists to deal with multicollinearity. Biased regression techniques such as ridge regression and principal component regression have been traditionally utilized for incorporating highly collinear independent variables into regression models. In particular, ridge regression is one of the most commonly used techniques for which the estimators can be described as follows ([Hastie et al., 2009](#)):

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (5)$$

where  $\lambda$  ( $\lambda \geq 0$ ) is the penalty parameter that determines the shrinkage. For instance, a larger  $\lambda$  determines a greater shrinkage of parameters toward zero ([Hastie et al., 2009](#)). The shrinkage can reduce the variance of ridge regression estimators when compared with LSE, with the cost of introducing some bias. From a multicollinearity standpoint, this reduction in variance is attractive, if not overly compensated by the increase in bias and mean square error ([Kutner et al., 2004](#)).

From an energetics standpoint, ridge regression and, more generally, biased regression techniques may provide an alternative to the univariate approach when PD and LD are highly correlated. This approach was explored by [Bernier et al. \(1987\)](#) who utilized ridge regression and principal components regression for the estimation of the energy efficiencies of protein and fat deposition in mice. The authors compared these two biased regression approaches with ordinary LSE using a database for which, according to the authors, PD and LD were highly collinear. The authors concluded that biased regression estimates produced biologically possible estimates of energy efficiencies, whereas the efficiencies estimated through ordinary LSE were biologically impossible and unstable. In particular, ridge regression estimators were less biologically likely than principal components regression and introduced some subjectivity in the analysis when compared with the LSEs ([Bernier et al., 1987](#)). The efficiencies of utilizing ME for PD were 0.37 and 0.39 for the two lines of mice presented by [Bernier et al. \(1987\)](#), whereas the efficiencies of LD were estimated at 0.51 and 0.69 for both lines of mice. It is important to note that the use of ridge regression has practical limitations, as most statistical methods. The penalty parameter has to be estimated, and

the use of statistical inference procedures developed for linear models may not apply directly (Kutner et al., 2004). More importantly, from an energetics standpoint, biased regression approaches address the multicollinearity but do not address the criticism that in equation (1), the role of dependent and independent variables is reversed (Koong, 1977).

### Multivariate Approach

The approach developed by Koong (1977) relied on the use of simultaneous equations. The author describes a previous attempt by Pullar and Webster (1974) to model protein and lipid deposition with a multivariate approach, but this initial approach assumed that the energy efficiencies for lean and fat gain were the same in rats (Pullar and Webster, 1974). To the best of our knowledge, Koong (1977) was the first study to describe protein and fat depositions as a set of simultaneous equations and that represented energy efficiencies of utilizing ME for each process.

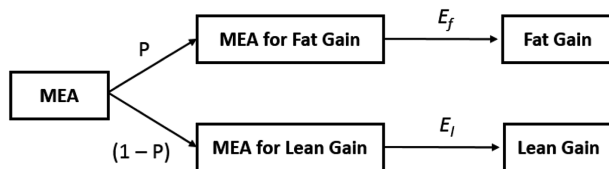
**The initial approach.** The framework proposed by Koong (1977) is described in Figure 2. In the first step, the metabolizable energy available for fat and protein deposition (MEA) is computed by removing the maintenance requirement (MR) or the  $ME_m$  from the ME intake:

$$MEA = MEI - MR \quad (6)$$

where MR denotes the maintenance requirement (or  $ME_m$ ). The author estimated MR with the following model:

$$BG/MBS = b_0 + b_1(MEI/MBS) \quad (7)$$

where BG is the total body energy gain (Mcal), MBS is the metabolic body size ( $kg^{0.75}$ ), and MEI is as before. Mathematically, the multivariate approach proposed by Koong (1977) can be described by the following equations:



**Figure 2.** Diagram to describe the Koong (1977) approach for estimating the energy efficiencies for fat and lean gain using simultaneous equations. Adapted from Koong (1977). The same notation from Koong (1977) was used to construct the diagram: MEA is the metabolizable energy available for fat and lean gain,  $P$  is the fraction of MEA utilized for fat synthesis, with  $(1 - P)$  as the remainder of MEA that is used for lean gain synthesis.  $E_f$  and  $E_l$  are the efficiencies with which the corresponding MEA fractions are utilized for fat and lean gain.

$$\begin{aligned} FG &= P \times E_f \times MEA \\ LG &= (1 - P) \times E_l \times MEA \end{aligned} \quad (8)$$

where FG is the energy gain in fat tissue, LG is the energy gain in lean tissue, and  $P$  is the fraction of MEA that is used for fat synthesis with  $(1 - P)$  as the remaining fraction of MEA that is used for protein synthesis. Furthermore, in the abovementioned notation,  $E_f$  and  $E_l$  were defined by Koong as the efficiencies of utilizing the corresponding fractions for MEA for fat and lean gain. It is important to note that we used the same notation as in Koong (1977), which is different than the more recent publications focused on energetics of growing animals (e.g., Strathe et al., 2010). For example, more recent publications define the responses as protein and lipid depositions (i.e., PD, LD) rather than fat and lean gain with the efficiencies of protein and lipid depositions defined as  $k_p$  and  $k_r$ .

The partitioning coefficient  $P$  is further modeled to allow the fraction of MEA that designated to fat vs. lean gain to change as a function of variables describing animals and dietary characteristics. In the Koong's (1977) model,  $P$  was described as a Michaelis–Menten type of equation with respect to the MEA:

$$P = \frac{MEA}{K + MEA} \quad (9)$$

where  $K$  is the parameter describing the level of MEA at which half of the MEA is used for fat synthesis. Therefore, the complete system of equations proposed by Koong can then be described by substituting equation (8) in equation (9):

$$FG = \frac{MEA}{K + MEA} \times E_f \times MEA \quad (10)$$

$$LG = \frac{K}{K + MEA} \times E_l \times MEA$$

Biologically, this set of equations describes the phenomenon that growing animals at low MEA supply, a greater proportion of MEA is utilized for LG, whereas at greater MEA supply, the proportion of MEA directed to FG increases and ultimately plateaus (Koong, 1977).

An application of the system was presented by Koong (1977) with a data set of growing mice of two lines. Equations were fit to data using a nonlinear least squares procedure based on a Nelder and Mead's (1965) type of algorithm that minimized the weighted sum of squares of residuals. The multivariate approach was compared with the univariate approach (equation (1)) that produced biologically impossible



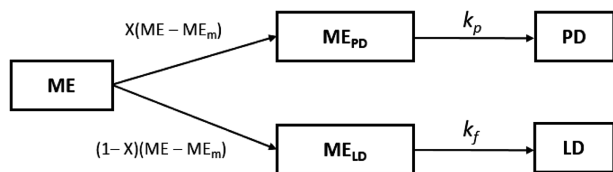
point estimates of  $E_p$ , that is, point estimates  $> 1$  and parameters with large standard errors. In particular, the univariate models estimated by [Koong \(1977\)](#) for two lines of growing mice were as follows: 1)  $MEI/MBS = 176.2 + 0.90 (\pm 1.05) \times FG/MBS + 7.11 (\pm 1.64) \times LG/MBS$  for the control line and 2)  $MEI/MBS = 170.4 + 1.40 (\pm 0.25) \times FG/MBS + 3.40 (\pm 0.92) \times LG/MBS$  for the fast-growing line. In this notation, MEI is the ME intake, FG is the fat gain, LG is the lean tissue gain, all expressed on a kcal basis, and MBS is the metabolic body size, that is,  $BW^{0.75}$ . The efficiencies in this univariate approach are given by the inverse of the parameters. Thus,  $E_f$  is 1.11 and 0.71 for the control line and fast-growing line, respectively. The  $E_l$  were 0.14 and 0.29 for the control and selected lines, respectively. When utilizing the multivariate approach,  $E_f$  and  $E_l$  were estimated at 0.36 and 0.17 for the control line and 0.57 and 0.39 for the selected line, respectively. It is important to note that these efficiencies estimated with the multivariate model are particularly smaller than what usually observed for monogastric animals, as reviewed by [Strathe et al. \(2012\)](#).

### Extensions of the Multivariate Approach

**An extension of the initial approach.** The approach proposed by [Koong \(1977\)](#) was extended by [van Milgen and Noblet \(1999\)](#) with the multivariate framework described in [Figure 3](#). Mathematically, this framework can be described with the following system of equations:

$$\begin{aligned} PD &= k_p X (ME - ME_m) \\ LP &= k_f (1 - X) (ME - ME_m) \end{aligned} \quad (11)$$

where ME is the metabolizable energy intake,  $ME_m$  is the metabolizable energy required for maintenance,  $X$  is the fraction of ME supplied above maintenance that is directed toward protein synthesis with the remainder fraction  $(1 - X)$  directed



**Figure 3.** Diagram showing the [van Milgen and Noblet's \(1999\)](#) multivariate approach for describing energy partitioning in pigs. Adapted from [van Milgen and Noblet \(1999\)](#). ME is the metabolizable energy intake,  $ME_m$  is the metabolizable energy required for maintenance,  $X$  is the fraction of ME supplied above maintenance ( $ME_m$ ) that is directed toward protein synthesis, with the remainder fraction  $(1 - X)$  directed toward lipid synthesis.  $ME_{PD}$  and  $ME_{LD}$  are the ME above maintenance that are directed toward protein and lipid synthesis,  $k_p$  and  $k_f$  are the efficiencies with which  $ME_{PD}$  and  $ME_{LD}$  are used for protein deposition (PD) and lipid deposition (LD), respectively.

toward lipid synthesis. It is important to note that we, again, utilize the notation as in the original manuscript, but the  $X$  fraction has a similar function as the  $P$  fraction proposed by [Koong \(1977\)](#). Similarly, the  $k_p$  and  $k_f$  efficiencies are denoted by  $E_l$  and  $E_f$  in the [Koong's](#) approach, respectively.

The fraction of MEA directed toward PD is modeled as a function of body weight:

$$X_i = c_i + d_i(BW - 20) \quad (12)$$

where  $c_i$  is the fraction of ME above maintenance directed toward PD at a BW of 20 kg and  $d_i$  is the change in the partitioning toward PD (vs. LD) with a unit change in BW. The index  $i$  describes the possibility of introducing between group variation into the parameters describing ME partitioning, for example, changes in the partitioning according to gender and genotype. This framework extends the previously developed multivariate model in many ways. First, it does not require an a priori estimate of the maintenance to solve the set of simultaneous equations. That is,  $ME_m$  is estimated simultaneously in equation (11) with the use of  $ME_m = a_i BW^b$  in equation (11) and the estimation of the parameters  $a_i$  (the  $ME_m$  requirement for the  $i$ th group, in [kcal or kJ/k BW]<sup>b</sup> per day) and  $b$  (the body weight metabolic exponent). Second, the authors formalized computations of total heat production (HP) and fasting heat production (FHP) using the model estimated parameters:

$$\begin{aligned} HP &= ME_m + (1 - k_p)X(ME - ME_m) \\ &\quad + (1 - k_f)(1 - X)(ME - ME_m) \end{aligned} \quad (13)$$

$$FHP = ME_m(k_p X + k_f(1 - X))$$

Moreover, [van Milgen and Noblet \(1999\)](#) comprehensively explored the interpretation of model parameters under different feeding conditions, proposed reparameterizations of the model for ad libitum feeding animals, and described a modification of the system for animals fed close to maintenance with zero energy retention ( $ME = ME_m$ ). The approach was demonstrated with the fitting of the multivariate model to a data set of growing pigs of seven combinations of sex and genotype. Overall, the approach from [van Milgen and Noblet \(1999\)](#) produced estimates of energy efficiencies that are consistent with biochemistry. Using their models, energy efficiencies varied between 0.58 and 0.60 for PD and 0.77 and 0.82 for LD. Interestingly, the authors reported that these estimates were similar to the ones computed with a univariate, factorial approach ([Noblet et al., 1999](#)) and attributed this

fact to the relatively low correlation between PD and LD in this data set (0.107). The authors have suggested, however, that the two approaches would probably produce different results if different feeding levels were used within the same genotype and BW, possibly increasing the correlation between PD and LD (van Milgen and Noblet, 1999).

The van Milgen and Noblet's (1999) approach has been successfully applied to a database of growing rainbow trout and Atlantic salmon by Azevedo et al. (2005). The overall correlation between PD and LD was relatively low in this database:  $-0.102$ . Estimates of energy efficiencies under the multivariate model ranged from 0.43 to 0.52 for PD and were centered at 0.81 for LD. When using a univariate model, estimates of the energy efficiency for LD for both rainbow trout and Atlantic Salmon were biologically unrealistic as they were estimated to be greater than one.

**The mixed model extension.** Both the approach of Koong (1977) and van Milgen and Noblet (1999) were developed and fitted under a multivariate, nonlinear regression approach. That is, the procedures do not directly consider the situation of grouped data, such as multiple animals from the same litter, and the possibility of multiple records collected on the same animal (Strathe et al., 2010). The approach of van Milgen and Noblet (1999) also ignores a potential covariance between the errors of PD and LD. In this context, Strathe et al. (2010) proposed a multivariate mixed-effects model to describe PD and LD. The authors argued that the proposed framework has the following advantages: 1) the possibility of describing PD and LD with nonlinear, diminishing return functions, such as the Gompertz functional form; 2) the extension of the multivariate approach to a mixed model framework that directly represents correlations on grouped data (Pinheiro and Bates, 2000) and allows the estimation of between animal variation; and 3) allows for a potential covariance between errors on PD and LD. In short, the mixed model framework proposed by Strathe et al. (2010) can be generally described as follows:

$$\begin{pmatrix} y_{i11} & y_{i12} \\ y_{i21} & y_{i22} \\ \vdots & \vdots \\ y_{im,1} & y_{im,2} \end{pmatrix} = \begin{pmatrix} f(\theta_i, ME_i, BW_i)_{i11} & f(\theta_i, ME_i, BW_i)_{i12} \\ f(\theta_i, ME_i, BW_i)_{i21} & f(\theta_i, ME_i, BW_i)_{i22} \\ \vdots & \vdots \\ f(\theta_i, ME_i, BW_i)_{im,1} & f(\theta_i, ME_i, BW_i)_{im,2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i11} & \varepsilon_{i12} \\ \varepsilon_{i21} & \varepsilon_{i22} \\ \vdots & \vdots \\ \varepsilon_{im,1} & \varepsilon_{im,2} \end{pmatrix} \quad (14)$$

where the  $n_i \times 2$  matrix on the left-hand side of equation (14) describes the response matrix for the  $i$ th animal (or, more generally, subject) with rows representing the  $n_i$  different observations and columns representing the two different responses: PD and LD. The first matrix on the right-hand side of the equation (14) has the functions to describe PD and LD and the last matrix on the right-hand side of equation (14) the matrix of errors.

The expectations of PD and LD are described by functions of ME intakes, BWs, and vector of animal-specific parameter  $\theta_i$  that contains both fixed and random effects (Pinheiro and Bates, 2000):  $\theta_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\boldsymbol{\gamma}_i$  where  $\boldsymbol{\beta}$  is a vector of fixed population parameters,  $\boldsymbol{\gamma}_i$  is a vector of random effects, and the matrices  $\mathbf{A}_i$  and  $\mathbf{B}_i$  are design matrices. Thus, random effects, describing between animal variation, may be incorporated into the model and are allowed to enter the model nonlinearly. The set of equations describing PD and LD expectations in equation (14) are given by:

$$\begin{aligned} \text{PD} &= \frac{\text{PD}_{\text{Potential}}}{1 + \frac{k \times \text{BW}}{\text{ME} - \text{ME}_m}}, \quad (\text{ME} - \text{ME}_m) > 0 \\ \text{LD} &= k_f \times \left( \text{ME} - \text{ME}_m - \frac{1}{k_p} \times \text{PD} \right) \end{aligned} \quad (15)$$

where  $\text{PD}_{\text{Potential}}$  is the parameter describing the potential PD,  $k$  is the saturation parameter, and  $k_p$  and  $k_f$  are the energy efficiencies of protein and fat deposition, as before. This approach determines the effect of limiting the energy supply on PD with a Michaelis–Menten type of equation. Biologically, under this model structure, PD approaches  $\text{PD}_{\text{Potential}}$  asymptotically as ME directed toward PD approaches infinity (Strathe et al., 2010).

The proposed model was fitted to two data sets of growing pigs, and when compared with the functional forms proposed by van Milgen and Noblet (1999), residuals suggested better ability of the simultaneous equations in describing the data sets. The models, however, proved to be challenging for achieving model convergence. For instance, in one of the data sets, a simultaneous estimate of maintenance (i.e.,  $\text{ME}_m = a_i \text{BW}^b$ ) could not be simultaneously obtained due to a failure of model to converge. The estimates of the efficiencies for PD ranged from 0.52 to 0.62 for the two data sets presented by the authors, whereas the efficiencies for LD ranged from 0.72 to 0.88.

**The Bayesian framework.** Parameter estimates from multivariate, nonlinear models are inherently harder to obtain than parameters obtained through LSE in equation (1). In fact, in the nonlinear mixed-effects

model, the marginal likelihood function of the data obtained by integrating out the random effects from the joint density of the data and random effects usually does not have a closed form expression (Pinheiro and Bates, 2000). Furthermore, the utilization of multivariate models can increase the number of model parameters, with the nonlinear model structure often reducing parameter identifiability and aggravating the issues with algorithm convergence. In this context, a Bayesian framework was proposed by Strathe et al. (2012) to estimate parameters in a multivariate model to describe energy intake and deposition in growing pigs. In particular, the authors argued that energetics experiments have been conducted for decades, and a considerable amount of a priori information is available in the literature. The combination of prior distributions and data likelihood in the construction of posterior density has the potential to assist with issues of parameter identifiability, especially in sparse data sets (Strathe et al., 2012). The Bayesian approach can be described as follows:

$$y_{ijk} | \beta_i, k_p, k_f, b, \Sigma \sim \text{MVN}(f(\beta_i, k_p, k_f, b, \text{BW}_{ij})_k, \Sigma) \quad (16)$$

where  $y_{ijk}$  is the  $k$ th observed response (MEI, PD, and LD, all expressed in MJ/d) for the  $i$ th group (in this example, groups represent animal's sex: barrow, boar, and gilt) at the  $j$ th BW ( $j = 1, \dots, n_i$ ),  $f(\bullet)_k$  is the expected of response  $k$  described by the set of multivariate equations (equation (17)), and  $\beta_i$  is the vector of parameters for the  $i$ th group. Furthermore,  $k_p$  and  $k_f$  are the partial efficiencies and  $b$  is the metabolic exponent parameter, as before. The variance covariance matrix  $\Sigma$  describes the residual variability in MEI, PD, and LD.

The multivariate approach from Strathe et al. (2012) extends the previous approaches by also treating ME intake as an additional response variable. In particular, ME intake is a response variable modeled as a function of BW, but it is also used as an explanatory variable for modeling LD. The set of equations describing the expected responses,  $f(\bullet)$ , are given by:

$$\begin{aligned} \text{MEI}_{ij} &= M_i(1 - e^{-k_i \times \text{BW}_{ij}}) \\ \text{PD}_{ij} &= \frac{\text{PD}_{\text{max}_i}}{\text{BW}_{\text{PDmax}_i}} \text{BW}_{ij} \times \log\left(\frac{\text{BW}_{\text{PDmax}_i} \times e}{\text{BW}_{ij}}\right) \\ \text{LD}_{ij} &= k_f \left( \text{MEI}_{ij} - a_i \times \text{BW}_{ij}^b - \frac{\text{PD}_{ij}}{k_p} \right) \end{aligned} \quad (17)$$

where  $\text{PD}_{\text{max}}$  is the parameter describing the pattern of maximum PD,  $\text{BW}_{\text{PDmax}}$  is the parameter

describing the BW at maximum rate of PD,  $M$  and  $k$  are the parameters describing the nonlinear relationship between MEI and BW, and all other parameters are defined as before. The Bayesian framework proposed by Strathe et al. (2012) requires the specification of prior distributions for all model parameters. For instance, Strathe et al. (2012) assumed that  $\beta_i \sim \text{MVN}(\mu, \mathbf{H})$  and  $\Sigma \sim \text{IW}(\mathbf{I}, 3)$ . The successful estimation of model parameters relied on utilizing literature information for constructing informative prior distributions, incorporated into the model through the mean vector  $\mu$  and variance-covariance matrix  $\mathbf{H}$ . An application of the model was presented with a data set of growing pigs for which parameters were estimated using a Markov Chain Monte Carlo approach (Strathe et al., 2012). The estimates of the efficiency of PD ranged from 0.58 to 0.62 for different models and sexes, whereas the efficiency of LD ranged from 0.73 to 0.78. It is important to note, however, that model convergence in the Bayesian framework can also be an issue and attention to model structure and form as well as specification of prior distributions will play a key role in model fitting and results.

## SUMMARY AND CONCLUSIONS

Energy efficiencies and maintenance parameters have been traditionally estimated with a linear regression model. Several studies have described the statistical issues associated with this approach, such as the reverse role of dependent and independent variables and a potential multicollinearity issue due to the high correlation between PD and LD. A first multivariate approach was developed by Koong (1977) who estimated the energy parameters using a set of simultaneous equations. This multivariate approach has been considerably extended in the past two decades with the complete characterization of model's biological interpretation under different feeding conditions, the simultaneous estimation of maintenance requirements, the extension of the model to a mixed-effects framework, and the implementation of a Bayesian framework for model fitting. Multivariate models are often harder to fit than linear regression models due to 1) larger number of parameters, 2) issues with parameter identifiability, 3) potential issues with overparameterization, and 4) overall lack of algorithm convergence. Therefore, with the recent availability of easy-to-use computer packages for model fitting (Plummer, 2003; Thomas et al., 2006; Carpenter et al., 2017), the use of a Bayesian framework seems as an attractive approach for

fitting multivariate models describing protein and lipid deposition.

### LITERATURE CITED

- Azevedo, A.P., J. van Milgen, S. Leeson, and D.P. Bureau. 2005. Comparing efficiency of metabolizable energy utilization by rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*) using factorial and multivariate approaches. *J. Anim. Sci.* 83:842–851. doi:10.2527/2005.834842x
- Bernier, J.F., C.C. Calvert, T.R. Famula, and R.L. Baldwin. 1987. Energetic efficiency of protein and fat deposition in mice with a major gene for rapid postweaning gain. *J. Nutr.* 117:539–548. doi:10.1093/jn/117.3.539
- Blaxter, K.L., and F.W. Wainman. 1966. The fasting metabolism of cattle. *Br. J. Nutr.* 20:103–111.
- Carpenter, B., A. Gelman, M.D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan: a probabilistic programming language. *J. Stat. Softw.* 76:1–32. doi:10.18637/jss.v076.i01
- Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction.* New York (NY): Springer.
- Koong, L.J. 1977. A new method for estimating energetic efficiencies. *J. Nutr.* 107:1724–1728. doi:10.1093/jn/107.9.1724
- Kutner, M., C. Nachtsheim, J. Neter, and W. Li. 2004. *Applied linear statistical models.* 5th ed. Chicago (IL): McGraw Hill/Irwin.
- Lofgreen, G.P., and W.N. Garrett. 1968. A system for expressing net energy requirements and feed values for growing and finishing beef cattle. *J. Anim. Sci.* 27:793–806.
- Moe, P.W., W.P. Flatt, and H.F. Tyrell. 1972. Net energy value of feeds for lactation. *J. Dairy Sci.* 55:945–958.
- Moraes, L.E., E. Kebreab, A.B. Strathe, J. France, J. Dijkstra, D.P. Casper, and J.G. Fadel. 2014. Bayesian analysis of energy balance data from growing cattle using parametric and non-parametric modelling. *Anim. Prod. Sci.* 54:2068–2081.
- Nelder, J.A., and R. Mead. 1965. A simplex method for function minimization. *Comp. J.* 7:308–313.
- Noblet, J., C. Karege, S. Dubois, and J. van Milgen. 1999. Metabolic utilization of energy and maintenance requirements in growing pigs: effects of sex and genotype. *J. Anim. Sci.* 77:1208–1216.
- Pinheiro, J., and D. M. Bates. 2000. *Mixed effects models in S and S-PLUS.* Statistics and computing. New York (NY): Springer.
- Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: K. Hornik, F. Leisch, and A. Zeileis, editors, *Proceedings of the 3rd International Workshop on Distributed Statistical Computing.* Technische Universitaet Wien, Vienna. (pp. 20–22).
- Pullar, J.D., and A.J. Webster. 1974. Heat loss and energy retention during growth in congenitally obese and lean rats. *Br. J. Nutr.* 31:377–392.
- Slinker, B.K., and S.A. Glantz. 1985. Multiple regression for physiological data analysis: the problem of multicollinearity. *Am. J. Physiol.* 249(1 Pt 2):R1–R12. doi:10.1152/ajpregu.1985.249.1.R1
- Strathe, A.B., A. Danfaer, A. Chwalibog, H. Sørensen, and E. Kebreab. 2010. A multivariate nonlinear mixed effects method for analyzing energy partitioning in growing pigs. *J. Anim. Sci.* 88:2361–2372. doi:10.2527/jas.2009-2065
- Strathe, A.B., H. Jørgensen, E. Kebreab, and A. Danfær. 2012. Bayesian simultaneous equation models for the analysis of energy intake and partitioning in growing pigs. *J. Agric. Sci.* 150:764–774.
- Thomas, A., B. O’Hara, U. Ligges, and S. Sturtz. 2006. Making BUGS open. *R News* 6:12–17.
- van Milgen, J., and J. Noblet. 1999. Energy partitioning in growing pigs: the use of a multivariate model as an alternative for the factorial analysis. *J. Anim. Sci.* 77: 2154–2162.