Predicting Average Feed Intake of Lactating Holstein Cows Fed Totally Mixed Rations

J. Fuentes-Pila,* M. Ibañez,* J. M. De Miguel,* and D. K. Beede†

*Department of Statistics and Management Sciences in Agriculture

Polytechnic University of Madrid

Ciudad Universitaria s/n, Madrid 28040, Spain

†Department of Animal Science

Michigan State University

East Lansing, MI 48824

ABSTRACT

A dry matter intake (DMI) prediction equation was estimated by using a data file that contained 124 treatment means collected from published studies. Animal factors considered for inclusion in the prediction model were body weight (BW) and its natural logarithm, $BW^{0.75}$, milk yield (MY) and its natural logarithm, milk fat and protein yields, month of lactation and its square, as well as its natural logarithm. The dietary factors considered were the percentages of neutral detergent fiber, acid detergent fiber, crude protein and hemicellulose in the ration dry matter together with the square of all these predictors. The multiple regression model selected by using the maximum R^2 method include both animal and dietary factors as independent variables. The accuracy of this DMI prediction equation was evaluated and compared with that of five other equations previously published by using three independent datasets also containing treatment means collected from literature. Even though the latest NRC equation was slightly more accurate than the equation proposed in this study with the three evaluation datasets, the latter can be used for some applications for which the NRC equation is not appropriate.

(**Key words:** feed intake, intake model, prediction equation, Holstein)

Abbreviation key: **FY** = milk fat yield, **MOL** = month of lactation, **MSPE** = mean square prediction error, **MY** = milk yield, **PY** = milk protein percentage, **RADF** = ration acid detergent fiber percentage, **RCP** = ration crude protein percentage, **RHEMI** = ration hemicellulose percentage, **RNDF** = ration neutral detergent fiber percentage, **RPE** = relative prediction error.

INTRODUCTION

Some method for predicting or forecasting feed intake is required in order to formulate TMR for dairy cows. Lack of accuracy in prediction may result in nutrient underfeeding or overfeeding affecting to animal performance, animal health or dairy farm environmental impact (NRC, 2001). Feed intake prediction inaccuracy may also limit the ability of different simulation and optimization techniques to improve the economic and technical efficiency of key operations in dairy farms such as feeding, breeding, or replacement (Fuentes-Pila, 1994).

Most models proposed for predicting voluntary feed intake of lactating dairy cows were developed empirically by applying multiple linear regression techniques (Ingvarsen, 1994; Forbes, 1995). Several studies (Fuentes-Pila et al., 1996; Roseler et al., 1997a) showed the lack of accuracy of existing feed intake prediction equations when evaluating them with independent datasets. The 2001 “Nutrient Requirements of Dairy Cattle” (NRC, 2001) recommended a DMI prediction equation based only on animal factors that were evaluated with treatment means from experiments published in the *Journal of Dairy Science* between 1988 and 1998. In the evaluation of the accuracy of prediction, this equation had a mean square prediction error of 3.31 kg^2/d and a bias of -0.27 kg/d (NRC, 2001).

In a meta-analysis study carried out at the Polytechnic University of Madrid, several DMI prediction equations for lactating Holstein cows fed TMR were developed by using multiple linear regression and incomplete principal component regression techniques (Fuentes-Pila, 1996). Because the parameters in these DMI prediction equations were estimated by using treatment means from experiments published in the *Journal of Dairy Science* between 1988 and 1995, the response variable was average DMI of lactating Holstein cow groups instead of individual cow DMI.

A first objective of this paper is to present the development and evaluation of one of the multiple linear regres-

Received February 15, 2002.

Accepted May 23, 2002.

Corresponding author: J. Fuente-Pila; e-mail: jfuentes@agricolas.upm.es.

sion prediction equations proposed in the above mentioned meta-analysis study. The intended use of this equation was the formulation of TMR with either corn silage or alfalfa (hay, silage, or haylage), or both as the forage base. These forages are widely used by dairy farmers in temperate climate regions of the world. For extending the use of this equation to the formulation of TMR including alfalfa-grass mixed forages, further evaluation would be needed. Likewise, because the developmental dataset did not include any diet containing grass forages, this equation shouldn't be used for predicting the intake of TMR including some grass forage before evaluating the effects of this extrapolation on prediction accuracy.

The new intake equation, shown in Table 1, included both dietary and animal factors as explanatory variables (Fuentes-Pila, 1996). For this reason, its incorporation into a least-cost ration formulation model will require the use of nonlinear programming techniques. These efforts are currently under way using a nonlinear programming stochastic model. A set of nonlinear restrictions, described in the Materials and Methods, have allowed us to simultaneously determine the DMI and the percentages of NDF, ADF, and CP in the least cost ration DM. The approach most commonly used in formulating least cost diets for dairy cows is to estimate nutrient requirements and feed intake before dietary ingredients are considered (NRC, 2001) in order to use these estimates as parameters in the linear programming model. But, as we discuss below, commercial mathematical programming software running on microcomputers can be used to formulate least cost rations by using nonlinear programming models that incorporate intake equations containing dietary factors.

The second objective of the current study was to evaluate the new intake equation with three independent datasets composed of treatment means from experiments published between 1985 and 1989 in the *Journal of Dairy Science*. Its prediction accuracy was compared with one of two previous intake equations based only on animal factors (Chase and Sniffen, 1985; Roseler et al., 1997b), two equations that include dietary factors as explanatory variables (Brown et al., 1977; Rayburn and Fox, 1993), and the equation recently recommended by the NRC (2001). All these equations are shown in Table 1.

MATERIALS AND METHODS

Dataset for Parameter Estimation

The data used for parameter estimation were collected from published studies including information on all the variables considered relevant to feed intake prediction. Treatment means were used as parameter esti-

Table 1. Previous DMI prediction equations evaluated in this study for lactating Holstein cows fed TMR.

Authors of the intake equation	Equation ¹
Brown et al. (1977) ²	$DMI = \exp(0.5198 + 0.000675 BW + 0.33922 \ln(MY) + 0.09927 FY - 0.000827 DIM + 0.14807 \ln(DIM) + 0.01800 (0.79 \times RADF - 0.000557 (0.79 \times RADF)^2))$
Chase and Sniffen (1985)	$DMI = 0.0185 BW + 0.305 FCM$
Rayburn and Fox (1993) ³	$DMI = 0.0117 BW + 0.281 FCM + 0.0749 DIM$, for $DIM < 84$ $DMI = 0.023 BW + 0.286 FCM + 0.0201 DIM - 0.0979 RNDF$, for $DIM \geq 84$
Roseler et al. (1997b) ⁴	$DMI = CF_{al} (8.4 + 0.006 BW + 12.2 PY)$, only for multiparous cows
NRC (2001)	$DMI = (-0.293 + 0.372 FCM + 0.0968 BW^{0.75}) \times (1 - \exp(-0.192 \times (WIM + 3.67)))$
Fuentes-Pila et al.'s new equation	$DMI = 548.0294 - 111.7076 \ln(BW) + 1.4024 BW^{0.75} - 4.4128 \ln(MY) + 3.7496 FY + 11.7684 PY + 3.8327 \ln(MOL) + 0.0043 RADF^2 + 0.7589 RNDF - 0.0101 RNDF^2 - 0.1367 RCP$

¹Definitions: MY = milk yield (kg/d); FY = milk fat yield (kg/d); PY = milk protein yield (kg/d); MOL = month of lactation, variable whose value is equal to 1 for the first month, 2 for the second one, 3 for the third one and 4 for the remaining months of lactation; RADF = ration acid detergent fiber (% of DM); RFND = ration NDF (% of DM); RCP = ration CP (% of DM); CF_{al} = multiplicative correction factor for early lactation which is 0.65 for the 1st week of lactation, 0.75 for the 2nd one, 0.81 for the 3rd, 0.87 for the 4th, 0.90 for the 5th, 0.92 for the 6th, 0.95 for the 7th, 0.96 for the 8th, 0.97 for the 9th one, 0.98 for the 10th, 0.99 from the 11th to the 16th, and 1.00 after the 17th week of lactation; PY = milk protein yield (kg/d); and WIM = weeks in milk.

²The crude fiber (CF) was substituted by the ADF by using the expression $CF = 0.79 \times ADF$ (Brown et al., 1977).

³The equation of Rayburn and Fox (1993) used in this study for predicting when DIM was greater than 84 was estimated by these authors with those treatments in which DIM was greater than 70.

⁴The values of CF_{al} used in this study were estimated for cows reaching peak milk yield in the second month of lactation. Instead of using BW at calving as explanatory variable, treatment mean BW was used. Although this equation was proposed by Roseler et al. (1997b) for multiparous cows, in this study it also was used for predicting the treatment mean DMI in dataset number 1 where treatments include both multiparous and primiparous cows.

mation data in some previous studies (Conrad et al., 1964; ARC, 1980; Rayburn and Fox, 1993). However, other researchers in this field used individual cow observations for estimating the feed intake (Ingvarlsen, 1994; Holter et al., 1997; Roseler et al., 1997b). Nonetheless, the meta-analysis approach based on using data from published trials is being applied widely in other fields such as medical research (Sutton et al., 2000) and applied social research (Lipsey and Wilson, 2000). We suggest that meta-analysis methods can be applied in dairy science research very cost effectively compared with the high cost of using experimental herds to carry out dairy research trials.

In this study, a dataset was compiled from published research with lactating Holstein cows in which cows were fed TMR in which the forage base was either corn silage, or alfalfa (hay, silage or haylage) or both; primiparous and multiparous cows were not included in the same experiment treatment; and, information on treatment period means was available for the following variables: DMI, BW, milk yield (**MY**), milk fat percentage, milk protein percentage, DIM, ration acid detergent fiber percentage (**RADF**), ration neutral detergent fiber percentage (**RNDF**), and ration CP percentage (**RCP**).

In the case of crossover-type experiments, only the control treatment was included in our compiled dataset. This approach was adopted to avoid having the cows and the diets used in a particular experiment overrepresented in our dataset. Recently, the Subcommittee of Dairy Cattle Nutrition also compiled a dataset from experiments published in the *Journal of Dairy Science* between 1992 and February 2000 for model evaluation, but they decided not to include any data from crossover type studies in their compiled dataset (NRC, 2001). We consider that both approaches are worthy of consideration but better in our opinion than including all data from crossover type studies as was done in a previous meta-analysis study (Rayburn and Fox, 1993). On the other hand, in the case of continuous lactation experiments where cows were assigned to each treatment at the beginning of the study and stayed on that same treatment until its completion, all the treatments were included in the dataset. The authors of the NRC meta-analysis study adopted a similar approach, but they only included data from experiments that lasted at least 6 wk (NRC, 2001). Again, we consider that both approaches can be considered acceptable.

The dataset used for parameter estimation consisted of 124 observations, representing treatment means. This dataset was compiled from 30 studies published in the *Journal of Dairy Science*. Table 2 summarizes the references of such studies, the number of treatment means from each study included in the dataset, and the forage base of the TMR fed in each study.

The variables considered as candidates to be included as independent variables in the regression equations were the following: BW, $BW^{0.75}$, lnBW, MY, lnMY, fat yield (**FY**), protein yield (**PY**), month of lactation (**MOL**; quantitative variable in which the value equals 1, 2, 3, and 4 for the first 4 mo of lactation and 4 for all remaining months of lactation), lnMOL, MOL^2 , RNDF, $RNDF^2$, RADF, $RADF^2$, RCP, RCP^2 , the ration hemicellulose percentage (**RHEMI**, defined as RNDF minus RADF), and $RHEMI^2$. Table 3 shows the mean, standard deviation, maximum, and minimum values for each one of the candidate variables representing linear effects.

These variables were chosen as candidates to be included in the regression model because they were included as independent variables in previous equations (Ingvarlsen, 1994; Forbes, 1995) and are reasonably obtainable in the field if the model has potential for practical application.

The FCM yield was not considered in this study because it is a linear combination of MY and FY. For this reason, there is exact collinearity when FCM yield, MY, and FY are included together in a linear regression model. Consequently, when stepwise procedures were used for selecting predictive models from sets of candidate variables including these three yield variables, results have differed depending upon the dataset used. Therefore, FCM yield was excluded from the model in the study carried out by Brown et al. (1977), whereas in the study of Yungblut et al. (1981a), the stepwise procedure excluded FY with one dataset and both FCM yield and FY with another one. However, because FCM yield was the variable most often used for estimating energy needs for lactation (INRA, 1989; NRC, 1989), it has been considered by some researchers the most adequate variable to account for the effects on DMI resulting from increased energy requirements of lactating dairy cows. Consequently, FCM yield is the yield variable used in several intake equations (Chase and Sniffen, 1985; Rayburn and Fox, 1993; Holter et al., 1997; NRC, 2001). But the recent NRC recommendations propose the use of FY, PY, and lactose yield, or of FY and PY if the lactose content of milk is not known, for estimating energy needs for lactation (NRC, 2001). Moreover, in the comprehensive database assembled by Roseler et al. (1997b), PY and MY had a slightly more positive correlation with DMI than did FCM yield. Therefore, in this study, the yield variables considered as candidates to be included in feed intake models were PY and FY together with MY and ln(MY).

We also preferred to use the variable MOL instead of DIM for preventing the possible association between DIM and some particular features of feed intake in some treatments that were not due to the stage of lactation.

Table 2. Studies selected from the *Journal of Dairy Science* for compiling the dataset used for estimating the feed intake prediction model.

First author/Publication year	Number of treatment means in the dataset	Forage in the TMR	Location	Condensed reference
Canale and Stokes (1988)	2	Corn silage (40%)	University of Maine	J. Dairy Sci. 71:373
Cassida et al. (1988)	2	Corn silage (50%)	Pennsylvania State University	J. Dairy Sci. 71:381
Gunderson et al. (1988)	1	Corn silage (19.4%), alfalfa hay (18.3%)	Colorado State University	J. Dairy Sci. 71:1204
Erdman et al. (1989)	8	Corn silage (40/60%)	University of Maryland	J. Dairy Sci. 72:1210
Jerred et al. (1990)	3	Alfalfa silage (45/64/84%)	University of Wisconsin	J. Dairy Sci. 73:2842
Austin et al. (1991)	2	Alfalfa hay (25%), corn silage (25%)	South Dakota State University	J. Dairy Sci. 74:3920
Cameron et al. (1991)	1	Alfalfa haylage (32.7%), corn silage (32.3%)	University of Illinois	J. Dairy Sci. 74:177
Faldet and Satter (1991)	3	Alfalfa silage (50%)	Agricultural Research Service, USDA Madison, Wisconsin	J. Dairy Sci. 74:3047
Grings et al. (1991)	4	Alfalfa hay (28.9%), alfalfa-grass silage (12.5%)	University of Idaho	J. Dairy Sci. 74:2580
Johnson and Combs (1991)	2	Alfalfa silage (31.2/37.0%)	University of Wisconsin	J. Dairy Sci. 74:933
Kim et al. (1991)	3	Corn silage (15.9/17.8%), alfalfa hay (25%)	South Dakota State University	J. Dairy Sci. 74:3891
Llamas-Llamas and Combs (1991)	1	Corn silage (25%), alfalfa silage (55.9%)	University of Wisconsin	J. Dairy Sci. 74:526
Broderick (1992)	14	Alfalfa silage (67.7 to 88.8%)	Agricultural Research Service, USDA Madison, Wisconsin	J. Dairy Sci. 75:174
Erickson et al. (1992)	8	Alfalfa grass silage (35%), corn silage (10%)	University of Illinois	J. Dairy Sci. 75:1078
Grings et al. (1992)	4	Alfalfa hay (21.3/21.6%), alfalfa silage (22.0/22.3%)	University of Idaho	J. Dairy Sci. 75:193
Johnson and Combs (1992)	1	Alfalfa silage (38%), corn silage (15%)	University of Wisconsin	J. Dairy Sci. 75:508
Shauff and Clark (1992)	1	Chopped alfalfa hay (15%), alfalfa haylage (22%), corn silage (13%) corn	Ohio State University	J. Dairy Sci. 75:2990
Shauff et al. (1992)	1	Alfalfa haylage (35%), corn silage (15%)	Ohio State University	J. Dairy Sci. 75:3003
Shauff et al. (1992)	1	Alfalfa haylage (45%), corn silage (5%)	University of Illinois	J. Dairy Sci. 75:1923
Zimmerman et al. (1992)	12	Alfalfa hay (16.5—48.3%)	North Carolina State University	J. Dairy Sci. 75:1954
Zimmerman et al. (1992)	16	Alfalfa hay (22.1—49.3%)	North Carolina State University	J. Dairy Sci. 75:1965
Broderick et al. (1993)	8	Alfalfa silage (25.5%), corn silage (30.1%)	Agricultural Research Service, USDA Madison, Wisconsin	J. Dairy Sci. 76:2266
Christensen et al. (1993)	4	Alfalfa hay (25%), corn silage (25%)	University of Illinois	J. Dairy Sci. 76:3490
Gaynor et al. (1994)	1	Corn silage (22.9%), alfalfa haylage (17.5%)	University of Maryland	J. Dairy Sci. 77:157
Drackley et al. (1994)	2	Alfalfa hay (32.5%), corn silage (17.5%)	University of Illinois	J. Dairy Sci. 77:1386
Johnson et al. (1994)	2	Alfalfa hay (7.6/7.5%), corn silage (39.2/38.6%)	Purdue University	J. Dairy Sci. 77:3676
Palmquist and Weiss (1994)	10	Alfalfa hay (25%), corn silage (25%)	Ohio State University	J. Dairy Sci. 77:1630
Broderick (1995)	2	Alfalfa hay (69.6/65.7%)	Agricultural Research Service, USDA Madison, Wisconsin	J. Dairy Sci. 78:320
Dado and Allen (1995)	1	Alfalfa silage (40.1%), corn silage (40.7%)	Michigan State University	J. Dairy Sci. 78:118
Grummer et al. (1995)	4	Alfalfa silage (52%)	University of Wisconsin	J. Dairy Sci. 78:172

Table 3. Means, standard deviations, and range of the variables considered for estimating the regression model in the developmental data set.

Variables	Mean	Standard deviation	Minimum	Maximum
DMI, kg/d	20.94	3.10	14.60	27.30
BW, kg	590.11	54.56	452.00	716.00
MY, kg/d	33.26	6.18	21.30	47.00
Milk fat, %	3.38	0.41	2.08	4.34
Milk protein, %	3.08	0.21	2.55	3.52
Milk fat yield, kg/d	1.11	0.19	0.61	1.62
Milk protein yield, kg/d	1.02	0.16	0.67	1.45
DIM	83.77	50.91	21.00	250.00
Month of lactation ¹	2.85	1.03	1.00	4.00
Ration ADF, % of DM	18.89	4.94	10.50	30.70
Ration NDF, % of DM	31.56	4.38	25.00	47.20
Ration CP, % of DM	17.97	2.54	13.30	27.10
Ration hemicellulose ² , % of DM	12.67	4.25	3.20	27.60

¹Month of lactation is a quantitative variable whose value is equal to 1 for the first month, 2 to the second month, 3 for the third month and 4 for the remaining months of lactation.

²Ration hemicellulose is defined as the difference between ration NDF and ration ADF.

Likewise, the variable MOL was assigned the numerical value 4 for months of lactation greater than 4 mo in order to obtain estimates of the effect of stage of lactation only during the first 3 mo of lactation. Some authors have considered that the effect of the stage of lactation on DMI is especially important during this portion of the lactation period (Roseler et al., 1997b; NRC, 2001).

The RHEMI was the only variable considered for inclusion in the multiple regression model that had not been included in previous DMI prediction models for lactating Holstein cows. However, in beef cattle, the silage hemicellulose content significantly affected voluntary silage intake and was included in some silage intake prediction equations (Rook and Gill, 1990).

Model Development

Using the option RSQUARE in PROC REG of SAS (1996), variable selection was carried out. The algorithm used by this procedure ensures finding the models with the highest R^2 values for each one of the possible sizes (1, 2, ..., k) of the model, where k is the total number of variables considered as candidates to be included in the regression model (Freund and Littell, 1991). Once this algorithm has identified the independent variable subsets with the highest R^2 values for each possible value of k, a criterion has to be adopted for choosing the model size. Among the different criteria proposed for variable selection (Hocking, 1976; Amemiya, 1980), the criteria considered in this study were: 1) following the recommendation of Hocking (1976) for choosing a predictive model, it was considered that the value of the C_p statistic (Mallows, 1973) should be lower than or equal to $p + 1$, where p

is the number of independent variables in the model; and 2) if the first criteria is satisfied, the J_p (Hocking, 1976), S_p (Hocking, 1976), PC (Amemiya, 1980), and AIC (Amemiya, 1980) statistics are computed and the model with the minimum values of these statistics is selected.

The values of the C_p , J_p , S_p , PC, and AIC statistics were computed with the PROC REG of the SAS (1996).

After selecting the independent variables to be included in the multiple linear regression model, the parameter estimation was performed by using the PROC REG of SAS (1996). The options RESIDUAL and INFLUENCE of this SAS procedure were used for outlier detection as recommended by Freund and Littell (1991).

The results of applying collinearity detection techniques and remedial measures for collinearity problems such as incomplete principal components regression or ridge regression will be shown in a future paper.

To study whether developing separate regression models for primiparous and multiparous cows or for diets exclusively based on alfalfa or corn silage, class variables were included in the selected model for testing heterogeneity of intercepts due to the effects of lactation number or forage type (Rawlings, 1988). In these models, location also was included as a qualitative variable. The procedure GLM of SAS (1996) was used for performing these multiple regression analyses with models including both quantitative and qualitative variables. If heterogeneity of intercepts was not detected, it was assumed that there was no heterogeneity of slopes because, in that case, there was no logic in expecting the two regressions to converge to a common intercept (Gujarati, 1992).

Table 4. Studies selected from the *Journal of Dairy Science* for compiling the three independent datasets used for evaluating the accuracy of the feed intake prediction model.

First author/Publication year	Independent dataset number	Number of treatment means in the dataset	Location	Condensed reference
De Peters and Taylor (1985)	# 1	2	University of California	J. Dairy Sci. 68:2027
Drackley and Schingoethe (1986)	# 1	2	South Dakota State University	J. Dairy Sci. 69:371
Yang et al. (1986)	# 1	2	South Dakota State University	J. Dairy Sci. 69:2348
Casper et al. (1987)	# 1	2	South Dakota State University	J. Dairy Sci. 70:321
Illg et al. (1987)	# 1	2	South Dakota State University	J. Dairy Sci. 70:620
Annexstad et al. (1987)	# 1	4	University of Minnesota	J. Dairy Sci. 70:814
Guillaume et al. (1987)	# 1	2	University of Minnesota	J. Dairy Sci. 70:2339
Casper and Schingoethe (1988)	# 1	2	South Dakota State University	J. Dairy Sci. 71:164
Schingoethe et al. (1988)	# 1	6	South Dakota State University	J. Dairy Sci. 71:173
Casper and Schingoethe (1989)	# 1	6	South Dakota State University	J. Dairy Sci. 72:928
Mäntysaari et al. (1989)	# 1	4	Cornell University	J. Dairy Sci. 72:2958
Broderick (1985)	# 2	3	US Dairy Forage Research Center	J. Dairy Sci. 68:3262
Shaver et al. (1986)	# 2	2	US Dairy Forage Research Center	J. Dairy Sci. 69:1545
Kung et al. (1987)	# 2	6	US Dairy Forage Research Center	J. Dairy Sci. 70:2069
Jaquette et al. (1988)	# 2	8	North Carolina State University	J. Dairy Sci. 71:2123
Woodford and Murphy (1988)	# 2	3	University of Illinois	J. Dairy Sci. 71:674
Kaiser and Combs (1989)	# 2	3	University of Wisconsin	J. Dairy Sci. 72:2301
Nakamura and Owen (1989)	# 2	3	University of Nebraska	J. Dairy Sci. 72:988
Oellermann et al. (1989)	# 2	2	Utah State University	J. Dairy Sci. 72:501
Weiss et al. (1989)	# 2	2	Ohio State University	J. Dairy Sci. 72:2308
Broderick (1985)	# 3	6	US Dairy Forage Research Center	J. Dairy Sci. 68:3262
Casper and Schingoethe (1986)	# 3	3	South Dakota State University	J. Dairy Sci. 69:1346
West et al. (1987)	# 3	4	Texas A & M University	J. Dairy Sci. 70:81
Atwell et al. (1988)	# 3	4	University of Illinois	J. Dairy Sci. 71:2689
Sharma and Erdman (1988)	# 3	4	University of Maryland	J. Dairy Sci. 71:2406
Sharma and Erdman (1988)	# 3	3	University of Maryland	J. Dairy Sci. 71:2670
Woodford and Murphy (1988)	# 3	3	University of Illinois	J. Dairy Sci. 71:687

Independent Datasets to Evaluate Accuracy of Prediction

Three independent datasets were compiled from experiments published between 1985 and 1989 in the *Journal of Dairy Science*. The experiments selected were those in which cows were fed TMR and treatment mean information was available for all variables considered in the model selection process.

Evaluation dataset number 1 includes 32 treatment means collected from 11 experiments in which treatments included primiparous and multiparous cows and the TMR were based either on alfalfa, corn silage, or both. Sources of data used for compiling dataset number 1 are shown in Table 4. Table 5 provides a description of this dataset.

Evaluation dataset number 2 is the result of selecting 32 treatments from 11 experiments published in which multiparous Holstein cows were fed TMR exclusively based on alfalfa. Sources of data also are shown in Table 4. The information in this dataset is summarized in Table 6.

In evaluation dataset number 3, 27 treatments were included. These treatments were from nine experiments, shown in Table 4, in which multiparous Holstein cows were fed TMR in which corn silage was the only

forage source. A description of dataset number 3 is shown in Table 7.

Model Evaluation

Validation of a regression model developed for predicting mean responses requires evaluating the accuracy of the fitted equation with independent datasets. The mean square prediction error (**MSPE**) as well as other statistics related to the MSPE are the measures most often used for evaluating the prediction accuracy of a model when being evaluated with independent datasets (Fuentes-Pila et al., 1996; Roseler et al., 1997a). Prior to the validation of a regression model with independent datasets, it is possible to estimate its prediction accuracy by using the statistic PRESS (Rawlings, 1988). This statistic is computed as follows:

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_{i,-i})^2$$

where y_i is the i th observation for the dependent variable, $\hat{y}_{i,-i}$ is the prediction of observation i using a model estimated without the i th observation but with all other observations, and n is the number of observations in the dataset used for parameter estimation. The value

Table 5. Description of the variables available for intake prediction in validation dataset number 1.

Variable	Mean	Standard deviation	Minimum	Maximum
DMI, kg/d	19.30	1.92	14.70	21.70
BW, kg	584.09	34.24	509.00	635.00
MY, kg/d	31.54	3.25	23.90	36.20
Milk fat, %	3.23	0.36	2.63	3.81
Milk protein, %	3.00	0.14	2.74	3.41
Milk fat yield, kg/d	1.01	0.11	0.77	1.28
Milk protein yield, kg/d	0.94	0.08	0.75	1.08
DIM	68.59	21.32	50.00	149.00
Month of lactation ¹	2.94	0.42	2.00	4.00
Ration ADF, % of DM	17.56	1.73	15.30	23.10
Ration NDF, % of DM	32.72	3.22	26.20	39.40
Ration CP, % of DM	15.84	0.79	14.80	17.70
Ration hemicellulose ² , % of DM	15.16	2.38	10.80	18.80

¹Month of lactation is a quantitative variable whose value is equal to 1 for the first month, 2 to the second month, 3 for the third month and 4 for the remaining months of lactation.

²Ration hemicellulose is defined as the difference between ration NDF and ration ADF.

Table 6. Description of the variables available for intake prediction in validation dataset number 2.

Variable	Mean	Standard deviation	Minimum	Maximum
DMI, kg/d	20.83	3.43	15.10	25.90
BW, kg	606.03	38.15	542.00	690.00
MY, kg/d	29.70	3.95	21.50	37.70
Milk fat, %	3.38	0.40	2.45	4.17
Milk protein, %	3.07	0.12	2.80	3.25
Milk fat yield, kg/d	1.00	0.14	0.73	1.29
Milk protein yield, kg/d	0.91	0.11	0.70	1.12
DIM	94.94	43.61	32.00	182.00
Month of lactation ¹	3.31	0.74	2.00	4.00
Ration ADF, % of DM	23.72	8.60	8.20	43.00
Ration NDF, % of DM	34.85	7.78	22.10	57.60
Ration CP, % of DM	18.25	2.83	13.80	23.80
Ration hemicellulose ² , % of DM	11.12	3.15	4.20	17.10

¹Month of lactation is a quantitative variable whose value is equal to 1 for the first month, 2 to the second month, 3 for the third month and 4 for the remaining months of lactation.

²Ration hemicellulose is defined as the difference between ration NDF and ration ADF.

Table 7. Description of the variables available for intake prediction in validation dataset number 3.

Variable	Mean	Standard deviation	Minimum	Maximum
DMI, kg/d	21.11	1.70	16.70	23.90
BW, kg	613.89	24.49	586.00	669.00
MY, kg/d	29.71	5.02	23.90	40.50
Milk fat, %	3.29	0.47	2.37	4.08
Milk protein, %	3.22	0.19	2.84	3.61
Milk fat yield, kg/d	0.96	0.12	0.73	1.17
Milk protein yield, kg/d	0.95	0.11	0.77	1.16
DIM	113.52	27.74	66.00	156.00
Month of lactation ¹	3.74	0.45	3.00	4.00
Ration ADF, % of DM	14.39	3.62	9.60	22.00
Ration NDF, % of DM	31.23	5.72	24.00	44.30
Ration CP, % of DM	16.10	2.11	13.70	22.30
Ration hemicellulose ² , % of DM	16.84	3.87	10.80	25.70

¹Month of lactation is a quantitative variable whose value is equal to 1 for the first month, 2 to the second month, 3 for the third month and 4 for the remaining months of lactation.

²Ration hemicellulose is defined as the difference between ration NDF and ration ADF.

of the statistic PRESS can be obtained with the options RESIDUAL or INFLUENCE of the PROC REG (SAS, 1996). The statistic PRESS divided by n can be considered as an estimate of the MSPE computed from the data used for parameter estimation. Although this MSPE estimate can be useful when independent datasets are not available, this statistic cannot detect lack of accuracy due to inadequacies of the sample used to estimate the regression model for providing a representation of the population studied (Rawlings, 1988).

Another statistic used in this study to evaluate the prediction accuracy of the equations developed is the relative prediction error (RPE) defined as the ratio between the positive root square of the MSPE and the value of $\bar{A}$. Some ad hoc criteria, based on the RPE values obtained with different independent datasets, have been proposed in order to decide whether the overall level of accuracy of a DMI prediction equation could be considered acceptable for its practical use (Fuentes-Pila et al., 1996). In this study, no ad hoc criteria are proposed. It was considered better to collect information from commercial dairy farm trials both on the prediction performance of the intake equations and on the economical consequences of its lack of accuracy before making any recommendation. Therefore, the objective of the accuracy evaluation carried out in this study is to provide a first estimation of the effectiveness of the developed equations for predicting the average DMI of lactating Holstein cow groups.

Comparison with Other DMI Prediction Equations

By using the same independent datasets, the accuracy of prediction of the intake equation developed in this study was compared with that of four previous equations (Brown et al., 1977; Chase and Sniffen, 1985; Rayburn and Fox, 1993; Roseler et al., 1997b) and of the equation recently recommended by NRC (2001). The four previous equations were selected because their predictions of the DMI of individual cows had the best levels of overall accuracy in an extensive evaluation (Fuentes-Pila, 1994). All five equations evaluated are shown in Table 1.

To predict DMI with the independent datasets used in this study, some modifications of the original equations were introduced (Table 1). In the case of the intake equation of Brown et al. (1977), the independent variable CF was estimated by using the same equation that Brown et al. (1977) used for estimating dietary CF from information on dietary ADF for some data in their developmental database. This equation was $CF = 0.79 \times ADF$. Another modification was the use of the past early lactation equation proposed by Rayburn and Fox (1993) after 84 DIM instead of after 70 DIM. Lastly, because

the treatment mean BW at calving was not reported in any of the published experiments selected for assembling the independent datasets, instead of using this explanatory variable, the BW treatment mean was used as a substitute in the intake equation proposed by Roseler et al. (1997b) for multiparous cows. This equation also was used for predicting the DMI treatment mean in independent dataset number 1, even though treatments in this dataset included primiparous and multiparous cows together.

Quantitative Methods for the Practical Use of Intake Equations Including Dietary Factors

As noted earlier, because of dietary factors included in the proposed feed intake equation, the incorporation of this equation into a least-cost ration formulation model requires adopting a nonlinear programming (NLP) formulation. For practical applications, the unavailability of both microcomputers fast enough and efficient enough to solve algorithms prevented the general use of decision support systems based on NLP models until the early 1990s. Currently, microcomputers much faster than the ones available a decade ago together with commercial algorithms such as MINOS (Murtagh and Saunders, 1998) can be used for solving NLP models for ration formulation in a very short time. State-of-the-art software development technology facilitates the integration of NLP models, the latest solving algorithms and databases in a user-friendly environment (De Miguel, 2000).

To use the new intake equation in the formulation of least cost TMR for management groups of lactating cows, an NLP model has been developed in the Polytechnic University of Madrid that incorporates this equation for predicting average feed intake. The model was written in the GAMS programming language after being formulated algebraically in summation notation and solved by using the MINOS algorithm integrated in the GAMS commercial software (Brooke et al., 1998). The model consists of the typical objective function and linear restrictions of least cost ration formulation models (Black and Hlubik, 1980) and some restrictions that allow the incorporation of the intake equation into the model. In this paper, only the latter restrictions are specified in order to illustrate how it is possible to simultaneously determine the DMI and the percentages of NDF, ADF, and CP in the least cost ration DM when solving the NLP problem.

The first of these restrictions is one defining the dry matter intake as:

$$\left(\sum_j^m DM_j \times AMOUNT_j \right) - DMI = 0$$

where DM_j is the parameter DM content of the feedstuff j expressed as proportion, $AMOUNT_j$ is the decision variable whose value represents the optimal amount in kilograms of the feedstuff j to be included in the least cost TMR, and DMI is a decision variable that represents the average amount of the formulated TMR that is needed to fulfill the needs of the animal considered.

For defining the contents of the least cost TMR in ADF, NDF, and CP, three nonlinear restrictions are used such as the following one:

$$\left(\sum_j^m DM_j \times NDF_j \times AMOUNT_j \right) - DMI \times RNDF = 0$$

where NDF_j is the parameter NDF content, expressed as proportion of DM, in the feedstuff j and $RNDF$ is a decision variable that represents the NDF content, expressed as a proportion of DM, in the least cost TMR. Because the restriction includes the product of two decision variables, DMI and $RNDF$, it is nonlinear. Analogously, we have two more nonlinear restrictions for defining the decision variables $RADF$ and RCP , which represent, respectively, the ADF and CP contents in the least cost TMR.

The intake equation is incorporated to the model as an additional nonlinear restriction that can be expressed as:

$$\begin{aligned} PDMI - 0.0043 \times 10,000 \times RADF^2 - 0.7589 \times 100 \\ \times RNDF + 0.0101 \times 10,000 \times RNDF^2 + 0.1367 \\ \times 100 \times RCP = 548.0294 - 111.7076 \times \ln(BW) \\ + 1.4024 \times BW^{0.75} - 4.4128 \times \ln(MY) + 3.7496 \\ \times FY + 11.7684 \times PY + 3.8327 \times \ln(MOL) \end{aligned}$$

By using intrinsic functions of the GAMS programming language, the result of the right hand side expression will be a constant whose value is function of the parameters BW, MY, FY, PY, and MOL, whose definitions are the same as the ones given in Table 1. In the left hand side, we have the decision variables, $PDMI$, the predicted DMI for the least cost TMR, $RADF$, $RNDF$, and RCP . Because the square of the decision variables $RADF$ and $RNDF$ have been included, this restriction is also nonlinear. At last, we have the following linear restriction:

$$DMI - PDMI \leq 0$$

This set of restrictions makes possible the use of the new intake equation for formulating least-cost TMR. Currently, this least cost ration formulation model is being validated.

Multiple regression analysis techniques also provide information that can be used to obtain confidence interval estimates of the expected mean value of the response variable for a particular choice of values for the independent variables (Rawlings, 1988). In the particular case treated in this study, such estimates are given by the following expression:

$$PDMI \pm t_{(a,v)} \sqrt{s^2 \sum_i^{p+1} \sum_j^{p+1} x_{ij} VAR_i VAR_j}$$

where $PDMI$ is the predicted value of the response variable, average DMI, for a particular set of values for the independent variables included in the multiple regression model; $t_{(a,v)}$ is the value of the Student's t -distribution with v degrees of freedom, being v equal to the degrees of freedom of the residual, that puts $a/2$ probability in each tail; s^2 is the error mean square; p is the number of independent variables included in the regression model; x_{ij} is the entry ij of the matrix $(\mathbf{X}'\mathbf{X})^{-1}$; VAR_i is either the value of the independent variable i th or 1 in the case of the intercept. These confidence intervals are provided by statistical software packages such as the SAS System (SAS, 1996), even though it is unusual to find published the values obtained for the matrix $(\mathbf{X}'\mathbf{X})^{-1}$. This information can be used for developing a stochastic formulation of the least cost formulation model where the decision variable DMI is restricted to be less than or equal to the lower limit of a confidence interval of the decision variable $PDMI$ for some convenient value of a . In other words, we'll use a more conservative estimate of the average DMI that takes into account the random nature of the response variable. This risk programming technique has been previously applied for taking into account the uncertainty on the nutrient content of feedstuffs when formulating least-cost rations (St-Pierre and Harvey, 1986).

Another application of these confidence interval estimates is using them for establishing statistical process control (SPC) limits for the actual average intake in order to detect the presence of special or assignable causes of variation depressing feed intake. Statistical process concepts and techniques such as control graphs (Box and Luceño, 1997) can be applied in commercial dairy farms in order to improve the performance of the feeding management process by identifying and eliminating assignable causes of variation. In dairies feeding TMR to groups of cows, the actual value of the average DMI can be estimated by weighing the amount of the TMR not consumed by the group of cows. These estimates can be obtained periodically in order to use average DMI as a quality indicator for detecting assignable causes of variation (for instance, some disorder such as

acidosis). The confidence interval estimates would be developed with data that represent experimental situations in which we could assume that nutritional disorders tend to be minimized. Currently, this application of the new intake equation is being implemented by programming spreadsheets to generate the control graphs for a particular set of values of the independent variables included in the equation.

RESULTS AND DISCUSSION

Model Selection

The results of applying the variable selection procedure described in the Materials and Methods section are summarized in Table 8. This table shows the independent variables included in the models with the highest R^2 value within five model size classes. The model size is the number p of independent variables included in the regression model. The five models are the ones with the lower sizes that satisfy the condition $C_p \leq p + 1$ following Hocking's (1976) recommendation for choosing a predictive model. Table 8 also includes the values of $p + 1$ and of the statistics C_p , J_p , S_p , PC , and AIC for each model. The model selected for further evaluation was the one with the lowest values of the statistics J_p , S_p , PC , and AIC . This model is the one in Table 8, including 10 independent variables.

The five models shown in Table 8 include as independent variables $\ln(BW)$, $BW^{0.75}$, FY , PY , and $\ln(MOL)$. Likewise, all of these models include either ADF , or ADF^2 or both, and either CP or CP^2 . The model selected also includes $\ln(MY)$, NDF , and NDF^2 as independent variables.

The use of $BW^{0.75}$ in other intake prediction models for dairy cows has been limited (ARC, 1980; NRC, 1989; NRC, 2001). Other DMI prediction equations for dairy cows have included BW instead of $BW^{0.75}$ (Ingvarstsen, 1994; Holter et al., 1997; Roseler et al., 1997b). Stepwise procedures eliminated $BW^{0.75}$ and included BW in the regression model of Brown et al. (1977), whereas $BW^{0.75}$ was included and BW did not enter into the preliminary models of Yungblut et al. (1981a). There is a tendency for excluding $BW^{0.75}$ from the set of candidate variables used for model selection when developing linear models to predict DMI of dairy cows (Ingvarstsen, 1994; Holter et al., 1997; Roseler et al., 1997b). The results obtained in this study suggest that, together with BW , both $BW^{0.75}$ and $\ln(BW)$ should be included in the set of candidate variables.

Regarding the yield variables, both the PY and FY were included in all the models shown in Table 8. These results are consistent with the idea that DMI is affected by the energy needs for lactation, which are, according to the latest NRC (2001) recommendations, mainly a

Table 8. Independent variables included in the models with the highest R^2 within five model size classes.

Model size (p)	Variables selected ^{1,2}	P + 1	C_p	Difference (p + 1) - C_p	J_p	S_p	PC	AIC	R^2	Adjusted R^2
9	$\ln(BW)$, $BW^{0.75}$, $\ln(MY)$, FY , PY , MOL^2 , $\ln(MOL)$, $RA DF^2$, RCP^2	10	9.030	0.970	2.269	0.0186	0.2376	101.5	0.7978	0.7819
10	$\ln(BW)$, $BW^{0.75}$, $\ln(MY)$, FY , PY , $\ln(MOL)$, $RA DF^2$, $RNDF$, $RNDF^2$, RCP	11	8.398	2.602	2.252	0.0185	0.2359	100.6	0.8025	0.7851
11	$\ln(BW)$, $BW^{0.75}$, $\ln(MY)$, FY , PY , MOL^2 , $\ln(MOL)$, $RA DF^2$, $RNDF$, $RNDF^2$, RCP^2	12	8.808	3.192	2.256	0.0185	0.2363	100.8	0.8054	0.7863
12	$\ln(BW)$, $BW^{0.75}$, MY , FY , PY , MOL^2 , $\ln(MOL)$, $RA DF$, $RNDF^2$, RCP^2 , $RHEMI$, $RHEMI^2$	13	9.221	3.779	2.260	0.0186	0.2367	101.0	0.8082	0.7865
13	$\ln(BW)$, $BW^{0.75}$, MY , FY , PY , MOL^2 , $\ln(MOL)$, $RA DF$, $RA DF^2$, $RNDF^2$, RCP^2 , $RHEMI$, $RHEMI^2$	14	10.41	3.595	2.279	0.0188	0.2388	102.0	0.8097	0.7872

¹Definitions: MY = milk yield (kg/d); FY = milk fat yield (kg/d); PY = milk protein yield (kg/d); MOL = month of lactation, variable whose value is equal to 1 for the first month, 2 to the second month, 3 for the third month and 4 for the remaining months of lactation; $RA DF$ = ration ADF (% of DM); $RNDF$ = ration NDF (% of DM); RCP = ration CP (% of DM); and, $RHEMI$ = ration hemicellulose (% of DM).

²The model with 10 variables (Model size (p) = 10) was selected for additional evaluation.

function of the PY and the FY. PY is included as independent variable in two recent intake predictions equations (Holter et al., 1997; Roseler et al., 1997b). The selection of either $\ln(\text{MY})$ or MY could account for other effects of milk production on DMI that could not be due only to increased energy requirements. The highest rates of milk production over the lactation are often associated with mobilization of body reserves (NRC, 1989). In a recent DMI prediction study, weekly BW change of individual cows accounted for 5 to 10% of the variation associated with DMI, it was correlated positively with DMI, and it was included in the intake prediction models selected by stepwise and maximum R^2 selection methods (Roseler et al., 1997b). However, average weekly BW change of a group of cows is a variable that has not been considered here as appropriate for intake prediction partially because it was not reported in the published experiments used for assembling the developmental database. But another reason was that, because positive and negative BW changes of individual cows tends to compensate each other when computing the average BW change, the latter variable was considered neither meaningful nor adequate for practical use.

As in previous studies, a nonlinear effect of stage of lactation separate from milk yield was found. However, no other intake equation for lactating dairy cows has included the ADF^2 simultaneously with NDF and the NDF^2 . This empirical result does not support the idea of considering NDF as a more appropriate explanatory variable for predicting DMI of ruminants (Mertens, 1987).

Lactation Number and Forage Type Effects

Regression analysis including class variables were used to study the convenience of developing separate models for primiparous and multiparous cows or for diets only including alfalfa, and diets that included corn silage with or without alfalfa. These two qualitative effects as well as the effect of experimental location were included in a covariance model as independent covariables in the regression model selected. The Type III sums of squares were used to determine whether or not these qualitative effects were statistically significant (SAS, 1996). The experimental location effect was significant ($P = 0.0005$), the lactation number effect was not ($P = 0.2813$), and the forage type effect was ($P = 0.0321$). These results suggest the potential need for developing separate models for each of the two forage types, that is, only alfalfa on the one hand and corn silage with or without alfalfa on the other hand.

Average Feed intake Prediction Equation

The multiple regression equation selected is the one shown in Table 1. The statistical information relative to this equation is summarized in Table 9. As the statistic *PRESS* divided by n can be considered as an estimate of the MSPE computed from the data used for parameter estimation, estimates of the MSPE and the RPE were computed from the value obtained for the statistic *PRESS*. The value of this estimated MSPE was 2.27 kg^2/d .

Because the data used for estimating this regression equation were treatment means, its goodness of fit cannot be compared with the one of previous regression equations estimated by using individual cow observations. The R^2 values reported for the most recent regression equations estimated by using treatment means from the literature (Rayburn and Fox, 1993) were unusually high (0.99). It seems that these R^2 values were obtained by using the BMDP Statistical Software (BMDP, 1988) for multiple regression models with the intercept forced to zero. When the intercept is forced to zero, the R^2 value is redefined both in BMDP (1988) and in SAS (Littell et al., 1991) as the sum of squared errors divided by the uncorrected total sum of squares instead of by the corrected sum of squares. It is not meaningful to compare this redefined R^2 value for regression models that does not include an intercept with the R^2 value for a regression equation including an intercept (Littell et al., 1991).

Although it could be expected that the goodness of fit of a regression equation estimated by using treatment means would be greater than one estimated by using individual observations, two recent equations of the latter type had greater R-squared values with a model of smaller size (Holter et al., 1997; Roseler et al., 1997b).

Before attempting a biological interpretation of the apparent relationships between the DMI and the different independent variables found in this multiple linear regression study, it is necessary to assess whether the regression coefficient estimates obtained are affected by collinearity problems. In the case of detecting that some of these estimates are affected by collinearity problems, it is possible that such estimates do not reflect the real influence of the explanatory variable on the response variable (Rawlings, 1988). It should be taken into account that interpretation of responses of dependent variables to individual independent variables in multiple regression equations can be misleading because of peculiarities of the data and spurious correlations between the dependent variable and some independent variables (Moore, 2000). For establishing cause and effect relationships and quantifying the effect of an explanatory variable on a response variable, de-

Table 9. Parameter estimates and regression analysis information for the regression model selected for additional evaluation.

Variables ¹	Parameter estimates	Estimated standard errors	Probability: $b_1 = 0$
Intercept	548.029420	160.590221	0.0009
ln(BW)	-111.707636	32.741938	0.0009
BW ^{0.75}	1.402392	0.375857	0.0003
ln(MY)	-4.412830	2.591754	0.0914
FY	3.749597	1.220683	0.0027
PY	11.768434	2.740033	0.0001
ln(MOL)	3.832746	0.363232	0.0001
RADF ²	0.004276	0.001111	0.0002
RNDF	0.758893	0.395301	0.0574
RNDF ²	-0.010063	0.005749	0.0828
RCP	-0.136654	0.059354	0.0231

Statistics related to the goodness of fit of the model¹:

$s = 1.4383$ kg/d C.V. = 6.8697% $R^2 = 0.8025$ adjusted $R^2 = 0.7851$

PRESS = 282.2244 $n = 124$ Mean DMI = 20.9365 kg/d

Estimated MSPE = 2.2760 kg²/d Estimated MPE = 1.5086 kg/d

Estimated RPE = 7.21%

¹Definitions: MY = milk yield (kg/d); FY = fat yield (kg/d); PY = protein yield (kg/d); MOL = month of lactation, variable whose value is equal to 1 for the first month, 2 to the second month, 3 for the third month and 4 for the remaining months of lactation; RADF = ration ADF (% of DM); RNDF = ration NDF (% of DM); RCP = ration CP (% of DM); s = square root of the error mean square (kg/d); CV = coefficient of variation (%); R^2 = coefficient of determination (%); n = sample size; $DMI_{average}$ = Average DMI; PRESS = predicted residual sum of square statistic (kg²/d); MSPE = mean square prediction error (kg²/d); EMP = mean prediction error, computed as the square root of the MSPE (kg/d); RPE = relative prediction error, defined as the ratio between the positive root square of the MSPE and the value of $DMI_{average}$.

signed experiments should be used instead of multiple regression studies as much as possible.

The s and CV values are related to the prediction accuracy of a regression equation. The intake equation developed in this study had a value of s (1.44 kg/d) and a CV value (6.87%), lower than the s values (1.79 and 1.66 kg/d) and the CV values (10 and 8%) corresponding to the equations also estimated with treatment means by Rayburn and Fox (1993). As expected, this s value also was smaller than the s values (between 1.48 and 1.71 kg/d) of the DMI prediction equations estimated using individual animal observations by Roseler et al. (1997b). The s values were not reported for the equations developed by Holter et al. (1997) and the CV values were not reported by Holter et al. (1997) or Roseler et al. (1997b). In any case, comparing the MSPE values obtained with a number of independent datasets is the best way of comparing the accuracy of different DMI prediction equations.

Evaluation of the Prediction Accuracy with Independent Datasets

In the Tables 10, 11, and 12, the results of evaluating the DMI prediction accuracy of the different feed intake equations evaluated in this study are presented. For the new DMI prediction equation in this study, the MSPE values obtained with each one of the three independent datasets were 1.91, 5.87, and 2.20 kg²/d, re-

spectively, whereas the RPE values were 7.16, 11.64, and 7.02%. With these three independent datasets, the latest equation recommended by the NRC (NRC, 2001) was the only model that consistently predicted DMI more accurately according to the MSPE values obtained (1.45, 5.36, and 1.62 kg²/d). Because the DMI mean value in the developmental database was not reported for this intake equation (Rayburn and Fox, 1993; NRC, 2001), the RPE values could not be computed. The new equation recommended by NRC (2001) is a combination of the regression equation estimated by Rayburn and Fox (1993) from the DMI table in NRC (1989) and of an early lactation adjustment factor developed by Roseler et al. (1997b). The values in the DMI table are not observed values, but rather apparently are values proposed by the members of the Subcommittee on Dairy Cattle Nutrition that developed the recommendations for NRC (1989). These values were estimates of average DMI for different combinations of fat-corrected milk yield and metabolic weight based on the experience and knowledge of subcommittee members. The result of this DMI prediction accuracy comparison suggests that Bayesian approaches incorporating prior information based on experts' opinion should be considered in the development of more accurate DMI prediction equations.

But the major strength of the NRC (2001) equation also is its major limitation. Because the equation was fitted to a set of estimates based on the opinion of ex-

Table 10. Accuracy of the DMI prediction equations evaluated with the independent dataset number 1.

Authors of the intake equation	Accuracy measures ¹			Bias and variation measures ²							
	MSPE	MPE	RPE	$\bar{\delta}$	b	r ²	s _p ²	s _A ²	($\bar{A} - \bar{P}$) ²	s _p ² (1 - b) ²	s _A ² (1 - r ²)
Fuentes-Pila et al.'s new equation	1.91	1.38	7.16	-0.837	0.874	0.676	3.161	3.575	0.701	0.050	1.159
Brown et al. (1977)	2.65	1.63	8.44	1.294	1.465	0.807	1.344	3.575	1.674	0.291	0.688
Chase and Sniffen (1985)	1.25	1.12	5.78	0.091	1.225	0.674	1.605	3.575	0.008	0.082	1.164
Rayburn and Fox (1993)	13.58	3.69	19.09	-3.499	0.885	0.636	2.904	3.575	12.243	0.039	1.302
Roseler (1994)	1.55	1.24	6.44	-0.478	1.681	0.755	0.955	3.575	0.228	0.443	0.876
NRC (2001)	1.45	1.20	6.24	-0.720	1.278	0.776	1.698	3.575	0.518	0.132	0.799

¹Definitions: MSPE = mean square prediction error (kg²/d); MPE = mean prediction error (kg/d), defined as the positive square root of the MSPE; RPE = relative prediction error (%), defined as the MPE divided by the mean of the observed intake values in the independent datasets and expressed as a percentage.

²Definitions: $\bar{\delta}$ = mean bias (kg/d), defined as the difference between the mean of the observed values ($\bar{A}$) and the mean of the predicted values ($\bar{P}$), being the values in boldface those significantly different from zero ($P < 0.10$); b = regression coefficient for the regression of the observed values on the predicted values, being the values in boldface those significantly different from 1 ($P < 0.10$); r² = R²-value for the regression of the observed values on the predicted values, being the values in boldface those less than 0.50; s_p² = variance of the predicted values; s_A² = variance of the observed values.

Table 11. Accuracy of the DMI prediction equations evaluated with the independent dataset number 2.

Authors of the intake equation	Accuracy measures ¹			Bias and variation measures ²							
	MSPE	MPE	RPE	$\bar{\delta}$	b	r ²	s _p ²	s _A ²	($\bar{A} - \bar{P}$) ²	s _p ² (1 - b) ²	s _A ² (1 - r ²)
Fuentes-Pila et al.'s new equation	5.87	2.42	11.64	-0.807	0.880	0.551	8.099	11.378	0.651	0.116	5.109
Brown et al. (1977)	17.91	4.23	20.32	3.160	1.504	0.342	1.719	11.378	9.986	0.437	7.487
Chase and Sniffen (1985)	11.67	3.42	16.40	1.431	1.016	0.154	1.702	11.378	2.048	0.000	9.626
Rayburn and Fox (1993)	14.57	3.82	18.33	-0.822	0.246	0.246	4.948	11.378	0.676	2.813	11.082
Roseler (1994)	10.86	3.29	15.82	1.433	1.250	0.236	1.715	11.378	2.053	0.107	8.693
NRC (2001)	5.36	2.31	11.11	0.503	1.806	0.689	2.402	11.378	0.253	1.560	3.539

¹Definitions: MSPE = mean square prediction error (kg²/d); MPE = mean prediction error (kg/d), defined as the positive square root of the MSPE; RPE = relative prediction error (%), defined as the MPE divided by the mean of the observed intake values in the independent datasets and expressed as a percentage.

²Definitions: $\bar{\delta}$ = mean bias (kg/d), defined as the difference between the mean of the observed values ($\bar{A}$) and the mean of the predicted values ($\bar{P}$), being the values in boldface those significantly different from zero ($P < 0.10$); b = regression coefficient for the regression of the observed values on the predicted values, being the values in boldface those significantly different from 1 ($P < 0.10$); r² = R²-value for the regression of the observed values on the predicted values, being the values in boldface those less than 0.50; s_p² = variance of the predicted values; s_A² = variance of the observed values.

Table 12. Accuracy of the DMI prediction equations evaluated with the independent dataset number 3.

Authors of the intake equation	Accuracy measures ¹			Bias and variation measures ²							
	MSPE	MPE	RPE	$\bar{\delta}$	b	r ²	s _p ²	s _A ²	($\bar{A} - \bar{P}$) ²	s _p ² (1 - b) ²	s _A ² (1 - r ²)
Fuentes-Pila et al.'s new equation	2.20	1.48	7.02	0.011	0.583	0.435	3.571	2.792	0.000	0.620	1.577
Brown et al. (1977)	9.70	3.11	14.75	2.821	1.038	0.378	0.980	2.792	7.958	0.001	1.736
Chase and Sniffen (1985)	4.67	2.16	10.24	1.727	0.923	0.397	1.300	2.792	2.983	0.008	1.685
Rayburn and Fox (1993)	5.74	2.40	11.35	-0.660	0.234	0.093	4.722	2.792	0.436	2.771	2.534
Roseler (1994)	3.62	1.90	9.01	0.878	0.481	0.125	1.502	2.792	0.771	0.404	2.444
NRC (2001)	1.62	1.27	6.04	0.095	0.982	0.421	1.221	2.792	0.009	0.000	1.615

¹Definitions: MSPE = mean square prediction error (kg²/d); MPE = mean prediction error (kg/d), defined as the positive square root of the MSPE; RPE = relative prediction error (%), defined as the MPE divided by the mean of the observed intake values in the independent datasets and expressed as a percentage.

²Definitions: $\bar{\delta}$ = mean bias (kg/d), defined as the difference between the mean of the observed values ($\bar{A}$) and the mean of the predicted values ($\bar{P}$), being the values in boldface those significantly different from zero ($P < 0.10$); b = regression coefficient for the regression of the observed values on the predicted values, being the values in boldface those significantly different from 1 ($P < 0.10$); r² = R²-value for the regression of the observed values on the predicted values, being the values in boldface those less than 0.50; s_p² = variance of the predicted values; s_A² = variance of the observed values.

perts and in which random variation in DMI was ignored, this regression model is not appropriate to estimate confidence intervals for the expected values of the response variable. As we have commented above, these confidence intervals have two important practical applications. The first one is used to formulate dairy rations by using chance-constrained programming. The other one is their use for establishing statistical process control limits for the average DMI of groups of cows in order to detect the presence special or assignable causes of variation depressing feed intake.

The two other equations (Chase and Sniffen, 1985; Roseler et al., 1997b) were more accurate with the independent dataset number 1, but they had greater MSPE values with the other two datasets. The RPE values for these two equations when evaluated with the independent dataset number 2 (Table 11) were notably higher than with our new intake equation. The two intake equations including dietary factors as independent variables (Brown et al., 1977; Rayburn and Fox, 1993) were much less accurate than our new equation when compared with the three independent datasets.

MSPE Components

Tables 10, 11, and 12 also show the values of the MSPE components as well as the values of the mean bias, the slope (b) of the regression of the actual DMI treatment means on the predicted ones, the square of the correlation coefficient (r^2) between the actual and the predicted DMI treatment means, and the population variances of the actual and predicted DMI treatment means, denoted s_A^2 and s_P^2 , respectively. For the intake equation developed in this study, the mean bias was significantly different from 0 with independent datasets number 1 and number 2, and the value of b was significantly different from 1 with the dataset number 3. The values of MSPE component due to mean bias were between 0.000 and 0.701 kg²/d, the values of the component due to line bias were between 0.050 and 0.620 kg²/d, and the values of the component due to random variation were between 1.159 and 5.109 kg²/d. The component due to random variation clearly accounted for the largest proportion of the MSPE.

The DMI prediction equation (NRC, 2001) that was most accurate at predicting the DMI with these three independent datasets had values of b significantly different from 1 with datasets number 1 and number 2. The mean bias for this equation was significantly different from 0 with dataset number 1. The values of MSPE component due to mean bias were between 0.009 and 0.518 kg²/d, the values of the component due to line bias were between 0.000 and 1.560 kg²/d, and the values of the component due to random variation were between

0.799 and 3.539 kg²/d. Because this equation included only three independent variables, the s_P^2 values were much lower than the ones for our new intake equation. It appears that the NRC (2001) equation could have a tendency to a greater line bias than our intake equation, but the prediction accuracy of these two equations should be evaluated with more sets of independent data in order to confirm this tendency. Line bias is a more serious problem than mean bias because it indicates structural problems of the model (Fuentes-Pila et al., 1996).

CONCLUSION

Our new intake equation tends to provide more accurate predictions of the average feed intake of lactating Holstein cows than all except one of the previous intake equations compared in this study. The exception is the recent intake equation proposed by NRC (2001) that is slightly more accurate, with the three independent datasets used for model evaluation. Despite additional information is needed to confirm this tendency, a first conclusion of this study is the greater accuracy and robustness of the NRC (2001) equation when predicting average DMI. Trials are being planned to evaluate the accuracy of the new intake equation and the NRC (2001) equation for predicting the average DMI of groups of lactating Holstein cows in commercial Spanish dairy farms.

In addition to this first conclusion, we consider that the major advantage of the new intake equation with respect to the equation recommended by the NRC (2001) is its ability to provide reliable confidence interval estimates for the expected values of the average DMI. These estimates can be incorporated into risk programming models for least-cost ration formulation or used for applying statistical process control techniques to the dairy feeding process.

Moreover, having available a DMI prediction equation that includes dietary factors as explanatory variables allows some effects on feed intake of increasing the proportion of forage included in the TMR to be accounted for. In the current economic environment faced by Spanish dairy farms, substantial improvements can be realized by maximizing the amount of high quality forage incorporated to dairy rations.

REFERENCES

- Agricultural Research Council. 1980. The nutrient requirements of ruminant livestock. Commonwealth Agricultural Bureaux, Farnham Royal.
- Amemiya, T. 1980. Selection of regressors. *Int. Econ. Rev.* 21:331–354.
- BMDP Statistical Software-SOLO User's Guide. 1988. BMDP Statistical Software, Inc., Los Angeles, CA.

- Black, J. R., and J. Hlubik. 1980. Basics of computerized linear programs for ration formulation. *J. Dairy Sci.* 63:1366–1378.
- Box, G., and A. Luceño. 1997. Statistical control by monitoring and feedback adjustment. John Wiley and Sons, Inc., New York.
- Brooke, A., D. Kendrick, A. Meeraus, and R. Raman. 1998. GAMS. A User's Guide. GAMS Development Corporation, Washington, DC.
- Brown, C. A., P. T. Chandler, and J. B. Holter. 1977. Development of predictive equations for milk yield and dry matter intake in lactating cows. *J. Dairy Sci.* 60:1739–1754.
- Chase, L. E., and C. J. Sniffen. 1985. Equations used in "ANALFEED" a VisiCal template. Mimeo, Cornell Univ. Dep. Anim. Sci., Ithaca, NY.
- Conrad, H. R., A. D. Pratt, and J. W. Hibbs. 1964. Regulation of feed intake in dairy cows. I. Change in importance of physical and physiological factors with increasing digestibility. *J. Dairy Sci.* 47:54–62.
- De Miguel, J. M. 2000. Un modelo de programación matemática y un sistema informático soporte para el análisis de políticas agrarias en cultivos herbáceos. Aplicación a una comarca española. [A mathematical programming model and a decision support system to analyze agricultural policies for arable crops. Application to a Spanish region]. Ph. D. Thesis, Polytechnic University of Madrid, Madrid, Spain.
- Forbes, J. M. 1995. The Voluntary Food Intake and Diet Selection of Farm Animals. 2nd ed. Commonwealth Agricultural Bureaux, Slough, UK.
- Freund, R. J., and R. C. Littell. 1991. SAS System for Regression. 2nd ed. SAS Series in Statistical Applications. SAS Institute Inc., Cary, North Carolina.
- Fuentes-Pila, J. 1994. A model to predict feed intake of lactating dairy cows. M. S. Thesis, University of Florida, Gainesville.
- Fuentes-Pila, J. 1996. Predicción de la ingestión voluntaria de vacas Holstein en lactación alimentadas con dietas completas [Predicting the voluntary food intake of lactating Holstein cows fed total mixed rations]. Ph. D. Thesis, Polytechnic University of Madrid, Madrid, Spain.
- Fuentes-Pila, J., M. A. DeLorenzo, D. K. Beede, C. R. Staples, and J. B. Holter. 1996. Evaluation of equations based on animal factors for predicting intake of lactating Holstein cows. *J. Dairy Sci.* 79:1562–1571.
- Gujarati, D. N. 1995. Basic Econometrics. Third edition. McGraw-Hill, Inc., New York.
- Hocking, R. R. 1976. The analysis and selection of variables in linear regression. *Biometrics* 32:1–49.
- Holter, J. B., J. W. West, and M. L. McGilliard. 1997. Predicting ad libitum dry matter intake and yield of Holstein cows. *J. Dairy Sci.* 80:2188–2199.
- Ingvarsen, K. L. 1994. Models of voluntary food intake in cattle. *Livest. Prod. Sci.* 39:19–38.
- INRA (Institut National de la Recherche Agronomique). 1989. Ruminant Nutrition. Recommended Allowances and Feed Tables. R. Jarrige, ed. John Libbey Eurotext, London.
- Lipsey, M. W., and D. B. Wilson. 2000. Practical Meta-Analysis. Applied Social Research Methods Series, Volume 49. Sage Publications, Inc., Thousand Oaks, CA.
- Littell, R. C., R. J. Freund, and P. C. Spector. 1991. SAS System for Linear Models. 3rd ed. SAS Institute, Inc., Cary, NC.
- Mallows, C. L. 1973. Some comments on C_p . *Technometrics* 15:661–675.
- Mertens, D. R. 1987. Predicting intake and digestibility using mathematical models of ruminal function. *J. Anim. Sci.* 64:1548–1558.
- Moore, D. S. 2000. The Basic Practice of Statistics. W. H. Freeman and Company, New York.
- Murtagh, B. A., and M. A. Saunders. 1998. MINOS 5.4 User's Guide. Technical Report SOL 83-20R. Department of Operations Research, Stanford University, Stanford, CA.
- NRC. 1989. Nutrient Requirements of Dairy Cattle. 6th rev. ed. Natl. Acad. Sci., Washington, DC.
- NRC. 2001. Nutrient Requirements of Dairy Cattle. 7th rev. ed. Natl. Acad. Sci., Washington, DC.
- Rawlings, J. O. 1988. Applied Regression Analysis: A Research Tool. Wadsworth & Brooks/Cole, Pacific Grove, CA.
- Rayburn, E. B., and D. G. Fox. 1993. Variation in neutral detergent fiber intake of Holstein cows. *J. Dairy Sci.* 76:544–554.
- Rook, A. J., and M. Gill. 1990. Prediction of the voluntary intake of grass silages by beef cattle. 1. Linear regression analyses. *Anim. Prod.* 50:425–328.
- Roseler, D. K., D. G. Fox, A. N. Pell, and L. E. Chase. 1997a. Evaluation of alternative equation for prediction of intake for Holstein dairy cows. *J. Dairy Sci.* 80:864–877.
- Roseler, D. K., D. G. Fox, L. E. Chase, A. N. Pell, and W. C. Stone. 1997b. Development and evaluation of equations for prediction of feed intake for lactating Holstein dairy cows. *J. Dairy Sci.* 80:878–893.
- Rawlings, J. O. 1988. Applied Regression Analysis. A Research Tool. The Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.
- Sutton, A. J., K. R. Abrams, D. R. Jones, T. A. Sheldon, and F. Song. 2000. Methods for Meta-Analysis in Medical Research. John Wiley & Sons Ltd., New York.
- SAS Institute Inc., 1996. SAS/STAT User's Guide, release 6.03 edition. SAS Institute Inc., Cary, NC.
- St-Pierre, N. R., and W. R. Harvey. 1986. Uncertainty in composition of ingredients and optimal rate of success for a maximum profit total mixed ration. *J. Dairy Sci.* 69:3074–3086.
- Yungblut, D. H., J. B. Stone, G. K. Macleod, D. C. Grieve, and E. B. Burnside. 1981a. The development of feed intake prediction equations for lactating dairy cows. *Can. J. Anim. Sci.* 61:151–157.