MILK YIELD PREDICTION AT LATE LACTATION IN REPRODUCTIVE RABBIT DOES

ARNAU-BONACHERA A., CERVERA C., BLAS E., PASCUAL J.J.

Institute for Animal Science and Technology, Universitat Politècnica de València, Camino de Vera, s/n. 46071, Valencia, Spain.

Abstract: Separating kits and mother to determine milk yield at 4th wk of lactation (MY4) could have negative consequences on the training and development of young rabbits. In this work, a total of 313 lactation curves (28 d long), taken from 2 different trials, were used to fit regression models to estimate MY4. In both trials, females were subjected to a semi-intensive reproductive rhythm [insemination at 11 d post-partum (dpp) and weaning at 28 dpp], but diets, genetic types, parity order and day of controls were slightly different. The models included variables which, according to the bibliography, are related to milk yield and are often recorded in joint management (without separation of litters from mothers), such as litter size at weaning (LSW; both linear and quadratic), joint energy intake of doe plus litter at 4th wk of lactation (JEI; both linear and quadratic), perirenal fat thickness change (ΔPFTd) and milk yield at 3rd wk (MY3). The overlapping degree (OL) between current lactation and next pregnancy was included as a dummy variable, as well as their interactions with quantitative traits. To fit these models, 3 procedures were proposed to obtain accurate equations with biological meaning: Eq1, multiple linear regression (MLR) of data; Eq2, MLR with previous smoothing of sample distribution; and Eq3, MLR with previous smoothing and avoiding redundant samples and collinearities among variables. MY3 had a positive and relevant linear effect on MY4 for the 3 equations obtained (responsible for 39 to 50% of MY4 prediction). JEI had also a relevant role in MY4 prediction (28 to 61%), its positive effect being linear on Eq1, quadratic on Eq2 and both linear and quadratic on Eq3. ΔPFTd and LSW related traits were only included in Eq3, with a low relative weight, and OL inclusion did not improve prediction in any equation. Predicting MY4 was possible with the variables used, although certain precautions must be taken. Traditional MLR seems to predict central values properly, but extreme values poorly, whereas pre-treatment of data to smooth the dependent variable distribution appears to improve prediction of extreme values.

Key Words: rabbit, milk yield, prediction, energy intake, litter size.

INTRODUCTION

As in other mammals, milk yield is a strategy of the rabbit female to provide nutrients and energy addressed to ensuring survival and adequate development of the young offspring. Lactation is a crucial period, not only for survival, but also to achieve a suitable performance and maturity degree at weaning time (García-Quirós *et al.*, 2014). In fact, a negative correlation has been reported between weaning weight and post-weaning mortality (Fortun-Lamothe and Gidenne, 2000). In addition, lactation could be an exigent period for the rabbit females, with a large production of milk characterised by high nutrient concentration (Maertens *et al.*, 2006), which may affect their body condition and further reproduction (Pascual *et al.*, 2013).

On the other hand, not only the amount of milk is important, but also the moment, for both does and kits. There are several factors affecting milk yield at a specific lactation stage, such as genetic type (Savietto *et al.*, 2013, 2014; El Nagar *et al.*, 2014), diet (Pascual *et al.*, 2003), or the overlapping degree between current lactation and next pregnancy (Lebas *et al.*, 1972). Independently of these factors, when rabbits are strictly dependent on their mother

(0 to 21 d old) high milk yield is welcome. However, high milk yield at late lactation could reduce litter solid feed intake (Nizza *et al.*, 2009), leading to a more sudden weaning that would increase health risk index during the growing period (Quevedo *et al.*, 2006).

Newborn rabbits stay in the nest box until 17-20 d of life and rabbit females usually go into the nest only once a day (Hoy and Selzer, 2002). Milk yield control is quite easy during this period, weighing females before and after milking. From this moment on, kits are able to leave the nest, raising their solid feed intake progressively. Therefore, milk yield control requires litter and doe separation during late lactation. This separation could have consequences on the experimental procedure (increased cost and management) as well as in both female and litter behaviour. According to Gidenne and Lebas (2006), young rabbits pass from a single to a larger number of sucklings per day, combining liquid and solid meals, and begin caecotrophy behaviour at late lactation. In fact, Maertens *et al.* (1988) showed how mother-litter separation resulted in a lower weight during lactation and a higher mortality after weaning, while Faraldo *et al.* (2013) showed how this separation could worsen the learning of the young rabbits, which would reduce their solid feed intake before and after weaning.

Thus, development of a model that could allow milk yield prediction at late lactation would be interesting. To enhance model usability and robustness, the data population is usually set up promoting dependent variable variability, but its distribution is not frequently considered. For instance, the frequent normal distribution of data provides a large number of values around the mean, but very few at the extremes, which could lead to a poorer prediction for extreme values. On the contrary, with a uniform distribution in the whole range, the model would be expected to predict extreme and around mean values with similar accuracy.

Therefore, the aim of the present work was to develop a model to predict milk yield at late lactation with data available in normal management without female-litter separation. To choose the model, not only fitting characteristics, but also proper biological meaning and ability to reproduce differences between treatments were considered. This study will thus provide equations for milk yield prediction at late lactation and knowledge of the main factors affecting or driving it.

MATERIALS AND METHODS

Robust prediction model building requires the use of broad databases, adequate fitting procedures and further validation through the prediction of external data, as well as the evaluation of their suitability under experimental conditions. The following methodology was proposed to achieve this goal.

Databases

To increase the robustness of the models, data variability was promoted by including 2 data sets in the 3 regression equations (*set 1* and *set 2*). In addition, another different data set was used to validate the models proposed (*set 3*). The reproductive rhythm was semi-intensive in all trials (insemination at 11 d post-partum [dpp] and weaning at 28 dpp), but the diets, genetic types, parity order and control days were slightly different in some cases.

Set 1: Consisting of 47 rabbit does (crossbred from maternal lines A and V, Universidad Politècnica de València, UPV) controlled throughout 5 consecutive cycles (1st to 5th) with a total of 173 lactations. Does were fed with a commercial diet (10.0 MJ digestible energy (DE)/kg of dry matter (DM), 170 g crude protein (CP)/kg DM) and controlled at 18 and 28 dpp for perirenal fat thickness (PFT) as described by Pascual *et al.* (2004) and DE intake (cage consumption of females together with their litters). Litters were standardised at birth to 8-9 kits in primiparous or 10-12 kits in multiparous rabbit does. Litters remained in the nest boxes until 21 dpp, which were closed, except for the weekends. Once a day in the morning, rabbit does were allowed to nurse and daily milk yield at 3rd wk was measured, weighing females before and after milking (average of 4 d). At 4th wk, litters and does were allocated to different cages. Once a day in the morning, each doe was taken to the litter cage to nurse and daily milk yield at 4th wk was measured (average of 4 d).

Set 2: Consisting of 140 primiparous rabbit does from different genetic types of UPV reporting different lactation curves (Savietto *et al.*, 2013, 2014): 53 from LP line (GT1); 42 from generation 16 of line V (GT2); and 45 from generation 36 of line V (GT3). One half was fed a commercial diet (D1; 11.6 MJ DE/kg DM, 175 g CP/kg DM) and the other half was fed with a low energy fibrous diet (D2; 9.1 MJ DE/kg DM, 162 g CP/kg DM). Does were controlled

at 14 and 28 dpp to measure PFT, and from 21 to 28 dpp for DE intake (of both females and litters). Litters were standardised at birth to 9 kits and milk yield was measured as described in *set 1*.

Set 3: Consisting of 32 rabbit does coming from different genetic types (UPV) and parity order: 26 crossbred females in their 1st or 2nd lactation (crossbred from the maternal lines H and LP), 4 females from LP line (4th to 6th lactation), and 2 females from A line (4th and 5th lactation). All the females were fed a commercial diet (10.0 MJ DE/kg DM, 175 g CP/kg DM). Does were controlled at 17 and 28 dpp to measure PFT and DE intake (of both females and litters). Milk yield was measured as described in *set 1.*

Model traits

Daily milk yield during the 4th wk of lactation (MY4) was used as dependent variable. As independent variables, we included traits frequently related to milk yield (Maertens *et al.*, 2006) and measurable in management without female-litter separation:

a) Litter size at 28 d weaning (both linear [LSW] and quadratic [SQ{LSW}]), described as the main factor affecting milk yield (Lebas, 1987).

b) Daily energy intake recorded as joint (female+litter) energy intake during the 4th wk of lactation (both linear [JEI] and quadratic [SQ{JEI}]). Although correlated to LSW, JEI was included as the higher energy intake by the female usually corresponds with a higher milk yield (Xiccato, 1996) and could gather variability due to genetic and environmental factors (e.g. herd-year-season or diet).

c) Daily milk yield during 3rd wk of lactation (MY3) as a measure of the previous productive level of the doe.

d) PFT change at late lactation (ΔPFTd) as a measure of the possible antagonism between milk yield and body reserves accretion (Pascual *et al.*, 2002). This was calculated by obtaining the difference between PFT at weaning and the previous control (18, 14 or 17 dpp for dataset 1, 2 or 3, respectively) and dividing this difference by the time between weaning and the previous control (10, 14 or 11 d for dataset 1, 2 or 3, respectively).

The main descriptive statistics are presented in Table 1. Overlapping degree between current lactation and next pregnancy (OL, 1 for females that became pregnant while lactating and 0 for females that did not get pregnant) was also recorded.

Finally, diet, parity order and genetic type were also recorded as qualitative traits, but were not included in the equations. They were used to assess how the predicted values obtained by the different equations estimate the effects of each treatment (Casado *et al.*, 2006).

Regression model

The general model used for the regressions was:

$$MY4 = Intercept + \sum_{i=1}^{6} b_i \cdot X_i + b_7 \cdot OL + \sum_{i=1}^{6} b_i + 7 \cdot X_i \cdot b_7 \cdot OL + e$$

Where X_i are the 6 independent variables mentioned above (LSW, SQ[LSW], JEI, SQ[JEI], MYE and ΔPFTd). The model also included OL as a dummy variable (fixed effect) and its interactions with independent variables. The inclusion of OL allows parameterising of equations for pregnant and non-pregnant rabbit does and the interactions could help, where appropriate, to model different patterns of resources allocation for OL.

The general model considers many traits, some of which could be correlated or non-informative. Including all the variables in the equations could lead to overfitting and low parsimonious equations (Clarke *et al.*, 2009a). Therefore, only relevant variables that explain a considerable amount of variability were maintained in the equations. Consequently, 3 equations (including samples from sets 1 and 2) were tested to fit the general model, using a GLMSELECT procedure (SAS, 2009). A stepwise procedure was used to include variables in the equations and cross-validation as a standard tool for assessing model fit in a predictive accuracy sense (Clarke *et al.*, 2009a). *N*-fold was chosen as cross-validation type, splitting the calibration data set into 10 fractions (Zhang, 1993). A strict criterion of parsimony was applied to cope with the curse of dimensionality (Clarke *et al.*, 2009a) by avoiding including variables in the equations that did not improve prediction. During regression computing, a threshold of 3 was used

Table 1: Main descriptive statistics for the traits included in the models.

Variable	mean	min	max	SD	CV×100
All rabbits (n=346)					
MY4	208.8	65.0	350.0	49.9	23.9
LSW	8.85	4.00	12.00	1.39	15.67
SQ(LSW)	80.29	16.00	144.00	24.00	29.89
JEI[1]	4.08	1.00	6.07	0.81	19.77
SQ(JEI)[1]	17.30	1.00	36.89	6.60	38.12
MY3	250.5	65.0	405.0	60.5	24.2
ΔPFTd[1]	−0.0129	−0.2409	0.1179	0.0461	359
Set 1 (n=173)					
MY4	227.7	65.0	350.0	54.6	24.0
LSW	9.47	6.00	12.00	1.29	13.59
SQ(LSW)	91.40	36.00	144.00	23.66	25.89
JEI[1]	4.24	1.00	6.07	0.92	21.74
SQ(JEI)[1]	18.79	1.00	36.89	7.54	40.12
MY3	286.2	116.7	405.0	54.6	19.1
ΔPFTd[1]	−0.0094	−0.1147	0.1176	0.0367	390
Set 2 (n=140)					
MY4	192.2	81.7	278.3	36.7	19.1
LSW	8.08	4.00	10.00	1.06	13.16
SQ(LSW)	66.38	16.00	100.00	16.20	24.41
JEI[1]	3.89	2.28	5.88	0.63	16.22
SQ(JEI)[1]	15.50	5.18	34.59	5.10	32.89
MY3	212.4	65.0	293.3	43.3	20.4
ΔPFTd[1]	−0.0089	−0.1250	0.1179	0.0482	539
Set 3 (n=32)					
MY4	179.6	108.3	230.0	31.5	17.5
LSW	8.91	4.00	10.00	1.49	16.72
SQ(LSW)	81.47	16.00	100.00	23.41	28.73
JEI[1]	4.10	2.48	5.25	0.66	16.12
SQ(JEI)[1]	17.21	6.13	27.53	4.95	28.78
MY3	225.5	136.7	276.7	38.1	16.9
ΔPFTd[1]	−0.0487	−0.2409	0.0773	0.0654	134

[1] Calculated between the 2 control points recorded for each dataset (*set 1*: 18-28 dpp; *set 2*: 14-28 dpp for PFT and 21-28 dpp for energy intake; *set 3*: 17-28dpp). SD: standard deviation; CV: coefficient of variation; PFT: Perirenal fat thickness; MY4: Milk yield at 4th wk (g/d); LSW: Litter size at weaning; SQ(LSW): Square of LSW; JEI: Joint energy intake (MJ/d); SQ(JEI): Square of JEI (MJ^2/d^2); MY3: Milk yield at 3rd wk (g/d); ΔPFTd: PFT change (mm/d).

for studentised residuals to find outliers with a high residual value after predicting (e.g. bad fitting to the model). The 3 equations proposed were:

Equation 1 (Eq1). First, an outlier search for the independent variables was carried out. For the dependent variable, a threshold of (sample mean±2.5×sample standard deviation [SD]) was applied to eliminate extreme values, which are usually not sufficiently represented and could highly influence the regression. Finally, GLMSELECT was computed as previously mentioned.

Equation 2 (Eq2). Prediction of extreme values, when underrepresented, is usually poorer than for central ones. Increasing the relative weight of extreme values and decreasing it from central values in the equation would lead to a more uniform representation (smoothing of sample distribution) and a better prediction of extreme values. As in Eq1, outlier search and threshold for the dependent variable were applied. Afterwards, GLMSELECT was computed as in Eq1 but adding a weight variable in order to smooth sample distribution throughout the full range. This weight variable is not required for prediction because it only affects the regression coefficient estimates. To obtain it, the histogram of the dependent variable was split into 20 classes and the weight value was calculated as the ratio between the minimum number of values within a class and the number of values within its class.

Equation 3 (Eq3). It is possible to achieve more uniform prediction by using multivariate data techniques for robust and accurate equations, such as sample selection (Ferré and Rius, 1997; Dantas Filho *et al.*, 2004; Galvão *et al.*, 2005)

or principal component (PC) regression (Galeano Díaz *et al.*, 1997; Hemmateenejad *et al.*, 2007; Dinç *et al.*, 2008), based on principal component analysis (PCA). PCA transforms original variables into a lower number of new variables (PC, uncorrelated among them) that explain almost the whole variability of the originals. For this equation, a previous sample selection method was performed to try to improve the smoothing described in Eq2. First, the same threshold was also used for the dependent variable extreme values search. Afterwards, a PCA was computed without deleting any PC (Statgraphics, 2010). Each score from each sample was divided by the corresponding eigenvalue to obtain a dimensionless space with variance equal to one for each dimension. These transformations resulted in a Euclidean space where the distances are easily measured. Ten samples were randomly selected and the distance from each of these samples to the rest of the population was measured. All the neighbouring samples with a distance lower than one were deleted. The process (selection, measuring and deletion) was repeated again twice. Finally, another PCA was computed, where the score from each sample was divided by the new corresponding eigenvalue. The objective was to search for multivariate outliers by measuring distance of each sample to centre of the population (e.g. the centre of the population matches with the mean of the population and the origin of coordinates from the Euclidean space). The threshold was calculated according to Cuadras (2011) to compare means in multivariate inference, obtaining a value of 17.32 (P<0.01). Samples remaining in the set were used in the regression. The weight variable and GLMSELECT were computed as in Eq2, but using as independent variables those obtained in the 2nd PCA (variables were PCs and values were scores). This equation was characterised by a slightly smoothed distribution and non-redundant samples, (e.g. as redundant samples were removed from the set, the number of values per class in the histogram was lower and the smoothing of sample distribution lower).

Model suitability

To evaluate the different fitting suitability of the equations, root mean square error (RMSE) and R^2 (as recommended by Romero and Zúnica, 2005), as well as R^2 adjusted and Akaike information criterion (AIC, as recommended by Clarke *et al.*, 2009b), were determined. Means and SD of actual and predicted values for each equation were also obtained. As the equations could be used in further works to estimate the effects of different treatments (diet, genetic type, number of parturition etc.) on MY4, we studied the capacity of the equations to provide useful predicted values by computing a PROC GLM procedure (SAS, 2009) for the actual values and for the predicted values of each equation. The model included diet and genetic type as fixed effects. To evaluate the effect of parity order, a PROC MIXED was used (SAS, 2009) including parity order as fixed effect, animal as random effect and parturition number as repeated measure. The ability of the predicted values to reproduce the differences obtained with the actual values was evaluated.

Finally, the model was validated with an external data set (Set 3), calculating for each sample predicted values for each equation and comparing with actual values. Bias was evaluated as the mean difference between actual and predicted values. RMSE and R^2 were determined before and after bias correction.

RESULTS

The main statistics for the fitted equations are shown in Table 2. RMSE for Eq2 and Eq3 were slightly higher than those obtained for Eq1 (+4.7 and +9.7%, respectively). Actual and predicted means were similar between them for the 3 equations evaluated (max. difference of 1% for Eq2). Although there were no major differences, the highest R^2 was obtained for Eq1 (on av. +0.03 points compared to Eq2 and Eq3). However, AIC obtained for Eq3 was clearly lower (on av. 52% less compared to Eq1 and Eq2). Eq1 and Eq 2 had the lowest actual SD (5% less than Eq3), and Eq1 had the lowest predicted SD (18 and 26% less compared to Eq 2 and Eq3, respectively). Predicted range for Eq2 and Eq3 were close and with similar wideness to the actual (95 and 99% of actual range, respectively), whereas Eq1 predicted range was tighter than the actual (77% of actual range).

Table 3 shows the regression coefficients obtained for the 3 equations, only for those variables selected as relevant according to the procedure of each equation. When using centred and scaled data, intercept values were the predicted means. Parameters related to LSW were only included in Eq3, having only a relative weight of 14% on MY4 prediction. MY3 had a positive and relevant linear effect on MY4 for the 3 equations obtained (relative weight from 39 to 50%). JEI had also a relevant role in MY4 prediction (relative weight from 28 to 61%), its positive effect being linear on Eq1,

Table 2: Main statistics for the equations fitted to predict milk yield in the 4th wk of lactation.

Equation	Eq1	Eq2	Eq3
n	307	307	199
RMSE	29.9	31.3	32.8
R^2	0.590	0.549	0.565
R^2ADJ	0.588	0.547	0.563
AIC	2932	2961	1938
Mean			
Actual	213.1	213.1	213.7
Predicted	213.1	210.9	215.7
Range			
Actual	113.3-331.7	113.3-331.7	115.0-331.7
Predicted	137.1-304.6	128.0-336.2	111.6-327.4
Standard deviation			
Actual	46.7	46.7	49.3
Predicted	33.8	41.3	45.8

Eq1: Equation developed using standard general linear models and cross-validation as variable selection method. Eq2: Equation based on Eq1 but weighting the values of the dependent variable according to their frequency. Eq3: Equation based on Eq2 but using non-redundant samples and new transformed variables explaining almost all variability with a lower number of variables. RMSE: Root mean square error; R^2ADJ: Adjusted R-square statistic by the number of variables included in the model; AIC: Akaike information criterion.

quadratic on Eq2 and both linear and quadratic on Eq3. ΔPFTd was only included in the Eq3 with a low relative weight (9%). The inclusion of OL in the model did not improve prediction in any equation.

Table 4 shows the effect of parity order, diet and genetic type on MY4 when using actual and predicted MY4 data. For all least square means estimations, standard errors were slightly lower when using predicted rather than actual MY4 data. The effect of parity order and diet were significant ($P<0.001$) using both actual and predicted MY4 data.

Table 3: Regression coefficients of the variables selected as relevant for the equations developed to predict milk yield in the 4th wk of lactation.

	Equation		
	Eq1	Eq2	Eq3
Parameter estimates for centred and scaled data			
Intercept	213.10	210.91	215.65
b value:			
LSW			4.02
SQ(LSW)			4.42
JEI[1]	19.96		8.40
SQ(JEI) [1]		29.18	8.93
MY3	19.85	18.62	30.17
ΔPFTd[1]			5.26
Parameter estimates for raw data			
Intercept	28.93	57.84	11.32
b value:			
LSW			2.59
SQ(LSW)			0.165
JEI[1]	25.01		9.13
SQ(JEI) [1]		4.42	1.18
MY3	0.324	0.304	0.438
ΔPFTd[1]			8.27

Eq1: Equation developed using standard general linear models and cross-validation as variable selection method. Eq2: Equation based on Eq1 but weighting the values of the dependent variable according to their frequency. Eq3: Equation based on Eq2 but using non-redundant samples and new transformed variables explaining almost all variability with a lower number of variables (Shown coefficients for original variables). [1] between the 2 points recorded for each data set (*set 1*: 18-28 dpp; *set 2*: 14-28 dpp for PFT and 21-28 dpp for energy intake). LSW: Litter size at weaning; SQ(LSW): Square of LSW; JEI: Joint energy intake (MJ/d); SQ(JEI): Square of JEI (MJ²/d²); MY3: Milk yield at 3rd wk (g/d); ΔPFTd: PFT change (mm/d).

Table 4: Effect of the genetic type, diet and parity order in the milk yield at 4th wk of lactation (lsmeans, standard error in brackets) using the actual values or those obtained from the prediction equations of the set.

	Actual	Eq1	Eq2	Eq3
Parity order (*Set 1*, n=167)				
1	192.1(6.4)[a]	194.8(4.0)[a]	187.7(5.3)[a]	196.7(4.5)[a]
2	226.0(6.9)[b]	222.8(4.3)[b]	221.0(5.5)[b]	230.9(4.7)[b]
3	247.4(7.2)[c]	246.9(4.4)[cd]	251.9(5.8)[c]	256.9(4.9)[c]
4	267.9(7.8)[d]	254.4(4.9)[d]	260.0(6.3)[d]	269.5(5.5)[d]
5	235.1(7.4)[bc]	239.4(4.8)[c]	240.6(6.3)[c]	253.6(5.3)[c]
P-value	<0.0001	<0.0001	<0.0001	<0.0001
Diet (*Set 2*, n=139)				
D1	204.6(3.9)[b]	203.6(2.6)[b]	200.9(3.1)[b]	198.0(3.1)[b]
D2	180.5(4.0)[a]	186.0(2.6)[a]	178.0(3.1)[a]	178.9(3.1)[a]
P-value	<0.0001	<0.0001	<0.0001	<0.0001
Genetic type (*Set 2*, n=139)				
GT1	202.8(4.6)[b]	198.5(3.0)[b]	193.9(3.6)[a]	194.9(3.6)[b]
GT2	192.4(5.1)[ab]	189.3(3.3)[a]	184.3(4.0)[a]	181.0(4.0)[a]
GT3	182.3(4.9)[a]	196.7(3.2)[ab]	193.2(3.9)[a]	189.4(3.8)[ab]
P-value	0.0108	0.1091	0.1631	0.0371

Eq1: Equation developed using standard general linear models and cross-validation as variable selection method. Eq2: Equation based on Eq1 but weighting the values of the dependent variable according to their frequency. Eq3: Equation based on Eq2 but using non-redundant samples and new transformed variables explaining almost all variability with a lower number of variables. [a,b,c,d] Means at a same effect and column not sharing superscript differ significantly at *P*<0.05.

Actual contrast between D1 and D2 diets (+24 g/d; *P*<0.05) was similar to those obtained with predicted MY4 data (+18, +23 and +19 g/d for Eq1, Eq2 and Eq3, respectively). Something similar happened for parity order. For example, when comparing the most differing parities (4th-1st), contrasts were +75.8, +59.6, +72.3 and +72.8 for actual, Eq1, Eq2 and Eq3 data, respectively. However, while the effect of genetic type was significant when using actual or predicted MY4 values from Eq3 (*P*<0.05), no differences were found using predicted data from Eq1 or Eq2. Significant actual contrast between GT1 and GT3 females (+20.5 g/d; *P*<0.05) was not reproduced when using predicted data of any equation (always below +6 g/d; *P*>0.05).

A validation with external data was carried out to ensure a proper suitability of the equations (Table 5). Validation of Eq2 showed the lowest bias and RMSE as well as the highest R^2. Nevertheless, when bias was corrected for each equation, the 3 validations did not show great differences in their statistics.

DISCUSSION

As the aim of least squares is to minimise the sum of the residual squares, and Eq2 and Eq3 were designed from Eq1 trying to improve prediction of extreme values, Eq1 is expected to have the lowest RMSE and the highest R^2. However, the fact that actual and predicted means were similar, but predicted SD were lower and predicted range tighter than those obtained with the actual values, denotes that values under the mean were overestimated and those over the mean

Table 5: Validation statistics for the fitted equations to predict milk yield in the 4th wk of lactation (*Set 3;* n=32).

|		R^2		RMSE	
	Bias	Biased	Unbiased	Biased	Unbiased
Eq1	−16.22	0.277	0.551	26.36	20.78
Eq2	−10.18	0.418	0.525	23.65	21.35
Eq3	−17.84	0.126	0.457	28.99	22.85

Eq1: Equation developed using standard general linear models and cross-validation as variable selection method. Eq2: Equation based on Eq1 but weighting the values of the dependent variable according to their frequency. Eq3: Equation based on Eq2 but using non-redundant samples and new transformed variables explaining almost all variability with a lower number of variables. RMSE: Root mean square error. Bias: mean difference between actual and predicted values.
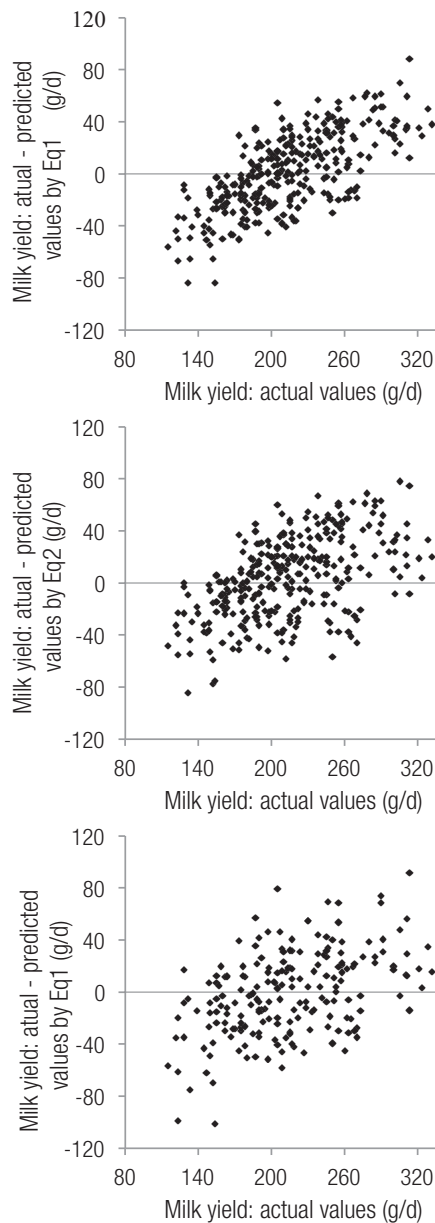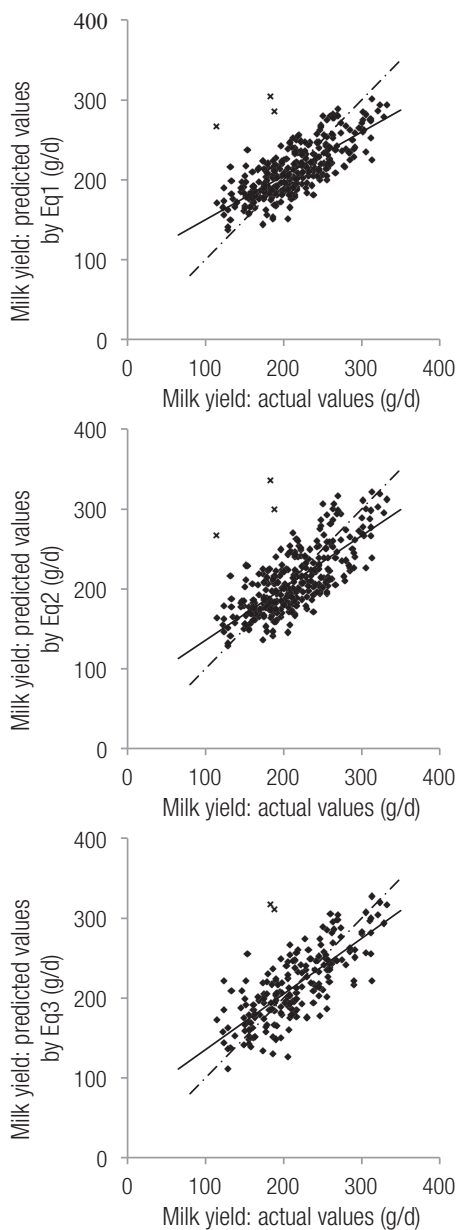
Figure 1: Calibration data set (*set1* and *2*). Relationship between predicted and actual milk yield for the equations developed. ◆ Samples used in the regression; × Regression outliers; —·— Actual=Predicted relationship; —— Least squares regression line for Predicted *vs.* Actual.

Figure 2: Calibration data set (*set1* and *2*). Relationship between residuals after predicted values for the equations developed and the actual values.
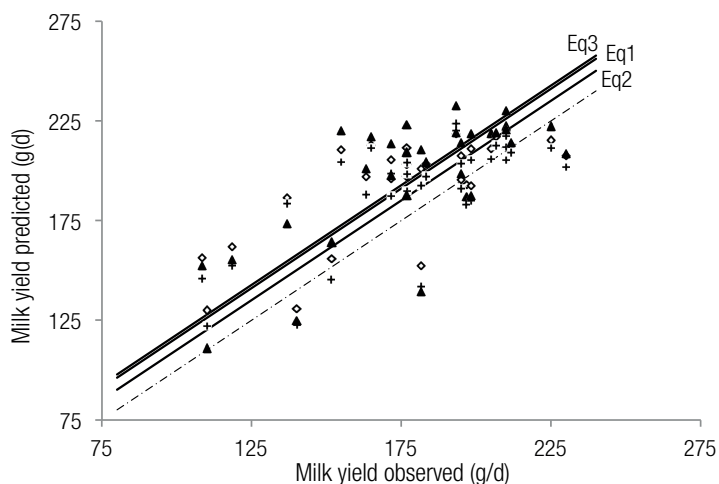
Figure 3: Validation data set (*Set 3*). Relationship between actual and predicted milk yield for each equation developed. ◇ Prediction of *Set 3* by Eq1; + Prediction of *Set 3* by Eq2; ▲ Prediction of *Set 3* by Eq3; – · – Actual=Predicted relationship; —— Actual=Predicted relationship corrected for average bias.

underestimated in general. As can be seen in Figure 1a, the slope between actual and predicted values with Eq1 was lower than 1 (0.55). This deviation was slightly improved when using Eq2 or Eq3 (slopes: 0.66 and 0.70, respectively), denoting a possible better prediction of extreme values. As a consequence, there was a higher linear correlation between residuals and actual values (Figure 2) for Eq1 (R^2=0.50) than for Eq2 and Eq3 (R^2=0.27 and 0.21, respectively). This higher correlation for Eq1 indicates that prediction error depends on the actual value and therefore would not be in agreement with the independence of the errors assumption on linear regression (Wilcox, 1997). The lowest value of AIC for Eq3 was mainly due to the great reduction in the number of samples used to fit the model (from 307 to 199). However, it also showed that the number of samples can be reduced (removing those redundant) without disturbing the fitting too much (e.g. when comparing R^2 or RMSE to those obtained for Eq2).

Interpreting the regression coefficients can give a biological meaning to the equations. The *b* values when using raw data indicate the change of MY4 from the intercept per unit of independent variable increase and these are the values that will be used later to predict MY4, while *b* values when using centred and scaled data indicate the change from the mean (intercept) of MY4 per unit of SD of independent variable increase, which allows us to compare coefficients among and within equations.

It was unexpected that the effect of LSW would be so scarcely relevant. Maertens *et al.* (2006) described litter size as the main factor affecting rabbit milk yield, supported by many authors (Lebas and Sardi, 1969; Torres *et al.*, 1979; Partridge and Allan, 1982; Pascual *et al.*, 1999). Two possible reasons are proposed, the first considering that as litter size rises, JEI and MY3 are also increased, so LSW effect could be confused or shared among them. The second considers that, as litter size was standardised at birth, it would have reduced the variability on LSW and consequently its impact on MY4. In any case, MY3 should reflect it. In fact, MY3 should take into account genetic and environmental factors (diet, litter size, temperature, etc.), considered as a predictor of the total daily amount of milk yield as suggested by Lebas (1976) or by the *k* value of the equation proposed in Casado *et al.* (2006). In this sense, it seems reasonable that the main factors affecting MY4 were the joint energy intake and previous milk yield. For instance, citing Lebas (1976), 58% of the variability among lactation curves could be explained by the first factor of a PCA which would be related to the total daily amount of milk yield. Consequently, females showing high milk yield at 3rd wk are expected to also have higher yield at 4th. Positive *b* values for JEI indicate that the higher the energy intake,

the higher the milk yields. This relationship is in agreement with that described by Xiccato (1996), where only linear effect was tested and observed.

Body condition change at late lactation (ΔPFTd) had no relevant effect on MY4 prediction. This could be due to the differences in time measurements among data sets or to the variability in factors affecting PFT changes included in the regression (overlaping degree, Pascual, 2010; parity order, Pascual *et al.*, 2013; genetic type and environmental conditions, Savietto *et al.*, 2013). As neither OL nor interactions with OL improved prediction of MY4, they were not included in any equation. This result would suggest no relevant change in nutrient partitioning between pregnant and non-pregnant females at 4th wk of lactation when inseminating at 11th dpp.

Direct comparison of coefficients between Eq3 and Eq1 or Eq2 is not possible, as the procedure to develop Eq3 forced all the variables to be in the model. PC used were independent in Eq3, but the original variables were still correlated (this fact did not allow separation of the effect of each independent variable on the dependent one). Regarding Eq1 and Eq2, both show positive *b* values for MY3 and for JEI, the latter linear in Eq1 and quadratic in Eq2. Although no life function has absolute priority over the other functions (Friggens and Van der Waaij, 2009), daily energy intake addressed to maintenance and survival could be considered close to fixed. Therefore, available energy for milk yield proportionally rises as energy intake increases, explaining the quadratic relationship between JEI and MY4 in the better prediction of extreme values with Eq2.

Regarding the possible use of predicted values to evaluate the effect of different treatments on milk yield, least square means obtained with the different equations were quite similar. Moreover, actual differences between treatments were in general higher than those obtained when using predicted values, as SD for actual values was higher than those for predicted (Table 2). However, improving prediction of extreme values (using Eq2 and Eq3) seems to increase SD of predicted values, as well as the standard errors and differences between treatments when using predicted values, their usability to reproduce treatment effects being improved.

The reproduction of genetic effects disturbed the discussion shown above. Although actual GT1-GT3 contrast was clearly significant (+21 g/d; *P*<0.05), the predicted values obtained from the equations were unable to reproduce it (from +0.7 to +5.5 g/d). This would be explained if RMSE equations were not absolutely random. For instance, there could be systematic errors depending on the genetic type, although these systematic errors must always be lower than RMSE, which biased least square means when using predicted values. Therefore, interpreting genetic results should be done with caution.

Validation of the proposed equations led to variable RMSE values, even unexpectedly lower than those obtained in the regressions (probably related to size and range of the validation set). Validation of Eq2 showed the lowest RMSE due to its lowest bias (as can be seen in Figure 3) and so the highest R^2. Taking into account that the mean of the validation set (180 g/d) was considerably lower than the mean of the calibration set (213 g/d), it seems reasonable that equations developed to improve prediction of extreme values presented lower bias and therefore better RMSE and R^2. This fact would highlight the relevance of a good fitting in the whole range (avoiding zones biased such as extremes). However, this improvement on bias was not observed when Eq3 was validated. Therefore, Eq2 again showed a better ability to predict MY4 values if actual values were not close to the mean.

CONCLUSIONS

Prediction of milk yield at 4th wk was possible with the variables proposed in this study, although certain precautions must be considered (especially taking into account the effect of genetic type). Further research would be required to reduce the prediction error. Main factors affecting milk yield at late lactation were energy intake and previous milk yield. The 3 models proposed showed similar ability to provide information that could be used to evaluate the effect of nutritional and reproductive treatments. Eq1 seems to predict central values properly, but extreme values poorly, whereas pre-treatment of data to smooth the dependent variable distribution seems to improve prediction of extreme values. Eq3 did not improve prediction and is more complex to apply (e.g. it requires the measure of body condition). Therefore, despite not presenting the highest R^2, Eq2 showed a good biological interpretation and seemed to be the most robust of those evaluated.

REFERENCES

Casado C., Piquer O., Cervera C., Pascual J.J. 2006. Modelling the lactation curve of rabbit does: Towards a model including fit suitability and biological interpretation. *Livest. Sci., 99: 39-49. doi:10.1016/j.livprodsci.2005.05.019*

Clarke B., Fokoué E., Zhang H.H. 2009a. Variability, Information and Prediction. *In: Principles and Theory for Data Mining and the Machine Learning. Springer New York, New York, USA, 1-43.*

Clarke B., Fokoué E., Zhang H.H. 2009b. Variable selection. *In: Principles and Theory for Data Mining and the Machine Learning. Springer New York, New York, USA, 569-671.*

Cuadras C.M. 2011. Inferencia Multivariante. *In: Nuevos Métodos de Análisis Multivariante. CMC Editions, Barcelona, Spain, 67-75.*

Dantas Filho H.A., Harrop Galvão R.K., Ugulino Araújo M.C., Da Silva E.C., Bezerra Saldanha T.C., José G.E., Pasquini C., Raimundo I.M., Rodrigues Rohwedder J.J. 2004. A strategy for selecting calibration samples for multivariate modelling. *Chemom. Intell. Lab. Syst., 72: 83-91. doi:10.1016/j.chemolab.2004.02.008*

Dinç E., Baleanu D., Ioele G., De Luca M., Ragno G. 2008. Multivariate analysis of paracetamol, propiphenazone, caffeine and thiamine in quaternary mixtures by PCR, PLS and ANN calibrations applied on wavelet transform data. *J. Pharm. Biomed. Anal., 48: 1471-1475. doi:10.1016/j.jpba.2008.09.035*

El Nagar A.G., Sánchez J.P., Ragab M., Mínguez C., Baselga M. 2014. Genetic comparison of milk production and composition in three maternal rabbit lines. *World Rabbit Sci., 22: 261-268. doi:10.4995/wrs.2014.1917*

Faraldo L., Ródenas L., Martínez-Paredes E. 2013. Variaciones del comportamiento alimentario de conejas y gazapos lactantes según su alojamiento. *In Proc.: XXXVIII Symposium de Cunicultura. 30-31 May, 2013. Zamora, Spain, 127-130.*

Ferré J., Rius F.X. 1997. Constructing D-optimal designs from a list of candidate samples. *TrAC - Trends Anal. Chem., 16: 70-73. doi:10.1016/S0165-9936(96)00084-2*

Fortun-Lamothe L., Gidenne T. 2000. The effects of size of suckled litter on intake behaviour, performance and health status of young and reproducing rabbits. *Ann. Zootech., 49: 517-529. doi:10.1051/animres:2000142*

Friggens N.C., Van der Waaij E.H. 2009. Modelling of Resources Allocation Patterns. *In: Rauw, W.M. (Ed.), Resources Allocation Theory Applied to Farm Animal Production. CABI Publishing, Wallingford, UK, 302-320.*

Galeano Díaz T., Guiberteau a., Ortíz Burguillos J.M., Salinas F. 1997. Comparison of chemometric methods: derivative ratio spectra and multivariate methods (CLS, PCR and PLS) for the resolution of ternary mixtures of the pesticides carbofuran Carbaryl and Phenamifos after their extraction into chloroform. *Analyst, 122*: 513-517. *doi:10.1039/a607955e*

Galvão R.K.H., Araujo M.C.U., José G.E., Pontes M.J.C., Silva E.C., Saldanha T.C.B. 2005. A method for calibration and validation subset partitioning. *Talanta, 67: 736-740. doi:10.1016/j.talanta.2005.03.025*

García-Quirós A., Arnau-Bonachera A., Penadés M., Cervera C., Martínez-Paredes E., Ródenas L., Selva L., Viana D., Corpa J.M., Pascual J.J. 2014. A robust rabbit line increases leucocyte counts at weaning and reduces mortality by digestive disorder during fattening. *Vet. Immunol. Immunopathol., 161: 123-131. doi:10.1016/j.vetimm.2014.07.005*

Hemmateenejad B., Akhond M., Samari F. 2007. A comparative study between PCR and PLS in simultaneous spectrophotometric determination of diphenylamine, aniline, and phenol: Effect of wavelength selection. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc., 67: 958-965. doi:10.1016/j.saa.2006.09.014*

Hoy St., Selzer D. 2002. Frequency and Time of Nursing in Wild and Domestic Rabbits Housed Outdoors in Free Range. *World Rabbit Sci., 10: 77-84. doi:10.4995/wrs.2002.479*

Lebas F., Sardi G. 1969. Alimentation lactée et croissance pondérale du lapin avant sevrage. *Ann. Zootech., 18: 197-208. doi:10.1051/animres:19690207*

Lebas F. 1976. Relations entre la forme de la curbe de lactation maternelle et la croissance des lapereaux. Etude sur 975 courbes individuelles. *In Proc.: 1ˢᵗ World Rabbit Congress, 1ˢᵗ April, 1976. Dijon, France, Communication 9.*

Lebas F. 1987. Influence de la taille de la portée et de la production laitière sur la quantité d'aliment ingérée par la lapine allaitante. *Reprod. Nutr. Dev., 27: 207-208. doi:10.1051/rnd:19870214*

Lebas F., Cousin M., Sardi G. 1972. Effet de la simultanéité de la lactation et de la gestation sur les performances laitières chez la lapine. *Ann. Zootech., 21: 129-131. doi:10.1051/animres:19720111*

Maertens L., Vermeulen A., De Groote G. 1988. Effect of post-partum breeding and pre-weaning litter management on the performances of hybrid does. *In Proc.: 4ᵗʰ World Rabbit Congress, 10-12 October, 1988. Budapest, Hungary, 141-149.*

Maertens L., Lebas F., Szendrő Zs. 2006. Rabbit milk: A review of quantity, quality and non-dietary affecting factors. *World Rabbit Sci., 14*: 205-230. doi:10.4995/wrs.2006.565

Nizza A., Stanco G., Di Meo C., Marongiu M.L., Taranto S., Cutrignelli M.I., Juliano L. 2009. Effect of pre-weaning solid feed and milk intake on caecal content characteristics and performance of rabbits around weaning. *Ital. J. Anim. Sci., 2: 95-101. doi:10.4081/ijas.2002.95*

Partridge G.G., Allan S.J. 1982. The effects of different intakes of crude protein on nitrogen utilization in the pregnant and lactating rabbit. *Anim. Prod., 35: 145-155. doi:10.1017/S000335610000091X*

Pascual J.J. 2010. The role of body condition on new feeding and breeding programmes for reproductive rabbit does. *In Proc.: 22ⁿᵈ Hungarian Conf. Rabbit Prod., 1-21.*

Pascual J.J., Cervera C., Blas E., Fernández-Carmona J. 1999. Effect of high fat diets on the performance, milk yield and milk composition of multiparous rabbit does. *Anim. Sci.*, *68*: 151-162.

Pascual J.J., Motta W., Cervera C., Quevedo F., Blas E., Fernández-Carmona J. 2002. Effect of dietary energy source on the performance and perirenal fat thickness evolution of primiparous rabbit does. *Anim. Sci., 75: 267-279.*

Pascual J.J., Cervera C., Blas E., Fernández-Carmona J. 2003. High-energy diets for reproductive rabbit does: effect of energy source. *Nutr. Abstr. Rev., 73: 27R-39R.*

Pascual J.J., Blanco J., Piquer O., Quevedo F., Cervera C. 2004. Ultrasound measurements of perirenal fat thickness to estimate the body condition of reproducing rabbit does in different physiological states. *World Rabbit Sci., 12: 7-21. doi:10.4995/wrs.2004.584*

Pascual J.J., Savietto D., Cervera C., Baselga M. 2013. Resources allocation in reproductive rabbit does: a review of feeding and genetic strategies for suitable performance. *World Rabbit Sci., 21: 123-144. doi:10.4995/wrs.2013.1236*

Quevedo F., Cervera C., Blas E., Baselga M., Pascual J.J. 2006. Long-term effect of selection for litter size and feeding programme on the performance of reproductive rabbit does 2. Lactation and growing period. *Anim. Sci., 82: 751-762. doi:10.1079/ASC200688*

Romero R., Zúnica L.R. 2005. Modelos de regresión. *In: Métodos Estadísticos En Ingeniería. Editorial Universitat Politècnica de València, Valencia, Spain, 297-334.*

SAS. 2009. User´s Guide (release 9.2). *Statistical Analysis System Institute Inc., Cary, NC.*

Savietto D., Cervera C., Blas E., Baselga M., Larsen T., Friggens N.C., Pascual J.J. 2013. Environmental sensitivity differs between rabbit lines selected for reproductive intensity and longevity. *Animal, 7: 1969-1977. doi:10.1017/S175173111300178X*

Savietto D., Cervera C., Ródenas L., Martínez-Paredes E., Baselga M., García-Diego F.J., Larsen T., Friggens N.C., Pascual J.J. 2014. Different resource allocation strategies result from selection for litter size at weaning in rabbit does. *Animal, 8: 618-628. http://dx.doi.org/10.1017/S1751731113002437*

Statgraphics. 2010. STATGRAPHICS® Centurion XVI User Manual. *StatPoint Technologies, Inc., USA.*

Torres A., Fraga M.J., Blas J.C. de. 1979. Milk production and mortality of New Zealand rabbits. *An. del Inst. Nac. Investig. Agrar. Prod. Anim., 25-30.*

Wilcox R. 1997. Introduction to the Robust Estimation and Hypothesis Testing. *Academic Press: San Diego, 14.*

Xiccato G. 1996. Nutrition of lactating does. *In: Lebas, F. (Ed.), Proc. 6th World Rabbit Congress, July, 1996. Toulouse, France, 29-50.*

Zhang P. 1993. Model selection via multifold cross validation. *Ann. Stat., 21: 299-313. doi:10.1214/aos/1176349027*