



HAL
open science

Comparison of forecast models of production of dairy cows combining animal and diet parameters

Thong Nguyen, Remy Fouchereau, Emmanuel Frenod, Christine Gerard,
Vincent Sincholle

► To cite this version:

Thong Nguyen, Remy Fouchereau, Emmanuel Frenod, Christine Gerard, Vincent Sincholle. Comparison of forecast models of production of dairy cows combining animal and diet parameters. Computers and Electronics in Agriculture, In press. hal-02358044v2

HAL Id: hal-02358044

<https://hal.science/hal-02358044v2>

Submitted on 22 Jan 2020 (v2), last revised 5 Feb 2020 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comparison of forecast models of production of dairy cows combining animal and diet parameters

Quoc Thong Nguyen^a, Rémy Fouchereau^b, Emmanuel Frénod^{a,b}, Christine Gerard^c, and Vincent Sincholle.^c

^a*Université de Bretagne Sud, Laboratoire de Mathématiques de Bretagne Atlantique, UMR CNRS 6205, Campus de Tohannic, Vannes, France*

^b*See-d, 6, rue Henri Becquerel - CP 101, 56038 Vannes Cedex, France*

^c*NEOVIA, France*

Abstract

We study the effect of nutritional diet characteristics on the lactating Holstein-Friesian dairy cows in Brittany, France from 36 individuals. An analysis of the relations between fat/protein content and milk yield was implemented for our dataset. The fat and protein production increase at a slower rate as milk yield increases. The importance of chemical composition on milk production is studied using the linear model. The data analysis confirms the importance of Starch, crude fiber, and protein which have a positive effect on milk production. This analysis also confirms the previous study on the effect of parity on the production. After that, the milk production forecasting is investigated using both linear models and machine learning approaches (support vector machine, random forest, neural network). We study the performance of multiple linear regression and machine learning-based models in both non-autoregressive and autoregressive cases at the individual level. The autoregressive models, which take into account the previously observed milk yield, have proven to significantly outperform the non-autoregressive approaches. Moreover, the computational cost of each approach is presented in the paper. While the random forest algorithm gives the best performance in both non-autoregressive and autoregressive approaches. The support vector machine algorithm gives a very close performance with a substantial less computing time. The support vector machine is shown to be the best com-

*Corresponding author

Email address: `quoc-thong.nguyen@univ-ubs.fr` (Quoc Thong Nguyen)

promise between accuracy and computational cost.

Keywords: Milk production forecasting, Dairy modeling, Autoregression, Smart farming

1. Introduction

Milk production forecasting of the dairy cow is an essential factor that is useful for the dairy farmers in management as well as health monitoring. In literature, many parametric models have been developed to model the lactation curve at the herd and individual level [1, 2, 3, 4, 5, 6]. Or the studies on extended lactation in dairy production [7, 8]. Recently, there are a number of modeling techniques on milk production forecasting that showed to obtain a highly accurate prediction with adaptability at the herd level [9, 10, 8]. The nonlinear autoregressive model with exogenous input using artificial neural networks introduced by Murphy et al. [9] shown to be most effective milk-production model.

On the other hand, understanding the effect of the nutritional diet on milk production and the quality of milk is not only helpful in financial planning but also in the production of other dairy products, such as yogurt, cheese, butter [11]. The importance of feed intake, diet on dairy cows was investigated in recent years. For example, the feed intake increases slowly at the beginning of lactation [12]; or the effects of dietary starch concentration on yield of milk and milk components were investigated by Boerman et al. [13].

In spite of that, not many studies are on individual cow level, and on the milk forecasting based on the nutrition for the small scale farms. Milk yield forecasting of each individual cow can be beneficial to many applications such as monitoring health conditions and disease detection, i.e. mastitis [14, 15]. Recently, Zhang et al. [16] conducted a study on the effect of parity weighting with the dataset in the south of Ireland; or Van Bebber et al. [17] applied Kalman Filter on monitoring dairy milk yields.

The subject of this study is to improve livestock farming, particularly milk production, by monitoring the performance in nutrition supplies. The first objective is to analyze the importance of the chemical composition of nutrition on the production and milk production monitoring of dairy cattle in Brittany, France. Secondly, we compare the performance of different types

31 of multiple linear regression and machine learning-based models for predic-
32 tion of production of the individual cow. The practicability and ability for
33 industrial applications are also discussed.

34 The paper is organized as follows. Section 2 is devoted to describe in detail
35 the content of our dataset and to present the composition analysis. Section 3
36 briefly recalls and analyzes the linear regression models and machine learning
37 algorithms. Section 4 focuses on the performance of the regression algorithms
38 on forecasting. The concluding remarks are given in Section 5.

39 2. Data description and composition analysis

40 2.1. Data description

41 The empirical data were collected from 36 lactating Holstein-Friesian
42 dairy cows in a research farm in Brittany, France, equipped with a robotic
43 milking system. For a ten months period (from December 2015 to September
44 2016), there are 7691 valid milking records collected. Each milking record
45 contains Daily Milk Yield (DMY), Day In Milk (DIM), parity information
46 (first, second, third onward lactation, see Tab. 1), number of milking per
47 day and the collective (corn silage, grass silage, wheat straw, soybean meal)
48 or individual (pelleted feed distributed through an automatic feeder) con-
49 sumption of diet components. Each cow is milked one to four times per day
50 by the robotic milking system, the cow can possibly be milked each time
51 it comes to the freestall for food. In this experiment, the amount of given
52 diets are changed every week. In this study, we are interested in the effect of
53 the diet on milk production forecasting. Particularly, the chemical compo-
54 sition studied in this paper are starch, crude fiber, Net Energy (NE) Unité
55 Fourragère Lait (UFL¹) and protein (PDIE²). Therefore, the consumption
56 of different diets was converted to these four chemical compositions. Table
57 2 presents the composition of each diet. It should be noted that, in Table
58 2, the consumption of the first eight diets (Corn silage, Grass silage, ..., Ni-
59 trogen supplement) is the same for 36 dairy cows at a specific week. On
60 the other hand, since the last four components (Production feed, ..., Liquid

¹which are respectively the units used in dairy production to estimate available energy and protein supply to dairy cows, estimated based on $1 \text{ UFL} = 1.7 \text{ Mcal}$, see [18].

²Protéines Digestibles dans l'Intestin limitantes par l'apport d'Énergie: true protein absorbable in the small intestine when rumen fermentable energy (organic matter) is limiting microbial protein synthesis in the rumen [19].

61 feed) in Table 2 are distributed by robot, which means the consumption of
 62 these four components varies according to the milk production level of each
 63 individual cow. Therefore, the consumption of each individual may differ at
 64 a specific week. In order to have a regular effect of each nutrient on milk
 65 production, we used the weekly data instead of the daily data. That means
 66 each data point is the average of seven days' observations. The statistical
 characteristics of the interesting variables are presented in Table 3.

Parity	number of cows
First lactation	20
Second lactation	13
Third onward lactation	3

Table 1: Number of individuals on each parity lactation.

	DM* content, %	Protein, g/kg of DM	Starch, g/kg of DM	Crude fiber, g/kg of DM	NE, UFL/kg of DM	PDIE, g/kg of DM
Corn silage	34.1	75	360	174	0.95	69
Grass silage	23.4	141	0	231	0.92	63
Fescue	88	93	0	222	0.76	82
Alfalfa hay	91.8	160	0	169	0.72	93
Fresh grass	18.3	167	0	217	0.94	90
Wheat straw	88	35	0	420	0.42	44
Ears corn	64	51	580	72	1.06	95
Nitrogen supplement	88	455	0	170	1.09	278
Production feed	88	273	114	14	1.17	205
Soluble nitrogen supplement	88	489	0	13	1.08	256
Ruminoprotected nitrogen supplement	88	443	0	13	1.08	273
Liquid feed	100	0	0	0	2.20	0

* Dry Matter

Table 2: Chemical composition of different diet.

67

68 2.2. Milk fat and protein composition analysis

69 In this section, we analyze the correlation between fat and protein content
 70 and milk yield with the collected data. The yield of cheese and butter mainly
 71 depend on milk fat and protein yield. A factor that impacts milk fat and
 72 protein concentration is milk yield [20]. It is well-known that, in daily rumi-
 73 nants, correlations among fat and protein content (g over 1 kilogram of milk
 74 yield) and milk yield are negative [21]. In our experiment, the reported cor-
 75 relation coefficients between milk yield and fat and protein content are -0.04
 76 and -0.21 , respectively. In our observed data, the fat and protein content

	Mean	SD ⁺	Min	Max
Starch (kg)	0.185	0.124	0.000	0.451
Crude fiber (kg)	0.426	0.190	0.080	0.966
PDIE (kg)	0.730	0.304	0.159	1.683
Net energy (UFL)	3.692	1.630	0.672	8.046
Parity	1.631	0.972	1	5
Milking per day	2.731	0.541	1	5

⁺ Standard deviation

Table 3: The statistical characteristics of the interested variables.

77 decrease as the milk yield increase, but not significant. As shown in Figures
78 1a and 1c, the fat and protein content visually decrease as milk yield increase
79 to 20 (kg/day). This phenomenon can be explained as at the beginning of
80 the lactation, the milk production increases more rapidly than the ability
81 of consumption of the cow. Moreover, when dairy cows produce more milk,
82 they consume more, especially water [22], but nutrition absorption cannot
83 change so intensively.

Some studies discovered that as milk yield increases, fat and protein syn-
thesis generally increases at a slower rate [23, 20]. This phenomenon can be
described by the allometric model:

$$y = ax^b$$

84 where y is fat or protein yield (g/day), x the milk yield (kg/day), and a and
85 b are equation coefficients. Parameter b represents a scaling factor describing
86 the effect of milk yield variation on its two main constituents. With $b = 1$,
87 milk yield shows a linear relationship with fat or protein yield whose content
88 in milk is equal to a ; if $b > 1$, fat or protein yield tends to increase more
89 proportionally than milk yield; and finally, if $b < 1$, fat or protein yield
90 increases at a slower rate than the milk yield.

91 In Figures 1b and 1d, the application of this model to data showed that
92 fat and protein synthesis varied proportionally to the output of milk with an
93 exponent 0.964 and 0.910 for milk fat and milk protein, respectively. Thus,
94 the higher the milk yield, the more cheese produced, even each additional
95 unit of milk results a lower increase in fat and protein. Moreover, from this
96 dataset, since the relationship between milk fat and milk yield has higher
97 variability than that between milk protein and milk yield (see Figure 1),

98 modification of milk composition by nutritional means should be easier to
achieve for fat than for protein.

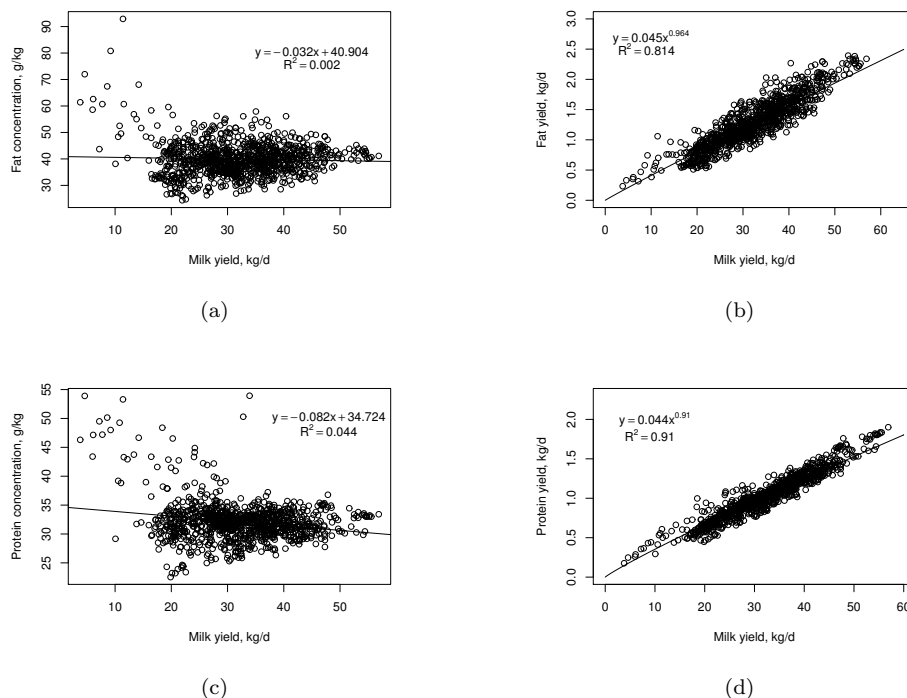


Figure 1: Relationships between milk yield and (a) milk fat yield, (b) milk fat concentration, (c) milk protein yield and (d) milk protein concentration.

99

100 3. Modelization

101 In this section, we present the linear models for analyzing the effect of the
102 features on milk production. Particularly, the fitting performance of three
103 linear regression methods (ridge, LASSO, elastic) is compared. In addition,
104 machine learning algorithms are introduced to predict milk production. The
105 multiple linear model is also used for forecasting. We compare the multiple
106 linear model with the machine learning approaches on milk prediction in the
107 next section.

108 *3.1. Multiple Linear Model*

109 A mixed linear model for milk yield observations is used. The model can
 110 be written as

$$y_{it} = \text{MPD} + \text{PAR} + \text{ST} + \text{CF} + \text{NE} + \text{PDIE} + f(t) + e_{it}, \quad (1)$$

where y_{it} = average of weekly milk yield of cow i at week t ; MPD = the fixed effect of Milking Per Day; PAR = fixed effect of parity; ST, CF, NE, PDIE are the fixed effects of the consumption of Starch (kg), Crude Fiber (kg), Net Energy (UFL), PDIE (kg), respectively; e_{it} = random residual error; they are assumed to be independent to each other. **The term $f(t)$ is the fixed function of week t based on the Ali and Schaeffer model [2], which is used to fit the average shape of the lactation curve. The Ali and Schaeffer model has been shown to be one of the most effective milk yield predictors [24, 16].** The model is written as:

$$f(i) = a_0 + a_1\gamma_t + a_2\gamma_t^2 + a_3\omega_t + a_4\omega_t^2,$$

where $\gamma = 7t/305$, $\omega = \ln(305/7t)$, and a_0, a_1, a_2, a_3, a_4 are regression coefficients. The coefficient a_0 is associated with the high of the general yield, a_1 and a_2 are associated with the increasing slope of the curve, a_3 and a_4 represent the decreasing slope of the curve. In matrix notation, the model can be given as

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e},$$

111 where \mathbf{y} is a $N \times 1$ vector of observed milk yield, \mathbf{b} is a $p \times 1$ vector of the
 112 regression coefficients, \mathbf{X} is an $N \times p$ incidence matrix, and \mathbf{e} is a $N \times 1$ vector
 113 of residual effects. Many regression methods have been developed to estimate
 114 the coefficients and improve the accuracy in prediction. In many problems,
 115 when the number of variables is too large, a selection model is needed to
 116 remove the less informative variables and reduce the computational cost. In
 117 some other cases, when the variables are highly correlated, another condition
 118 is required to prevent some variables from being poorly determined. In this
 119 study, we consider three common regression methods.

120 *Ridge regression*

Ridge regression is ideal if the features (the columns of \mathbf{X}) are highly related [25, 26]. In particular, it performs well with many features each having small effect and prevents coefficients with many correlated variables

from being poorly determined and exhibiting high variance. Ridge regression shrinks the coefficients of correlated features equally by penalizing. The ridge regression estimator solves the regression problem using L_2 norm penalized least squares:

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \{ \|\mathbf{y} - \mathbf{X}\mathbf{b}\|_2^2 + \lambda \|\mathbf{b}\|_2^2 \},$$

121 where $\|\mathbf{y} - \mathbf{X}\mathbf{b}\|_2^2 = \sum_{i=1}^n (y_i - \mathbf{x}_i^\top \mathbf{b})^2$ is the L_2 norm loss function, \mathbf{x}_i^\top is
 122 the i -th row of matrix \mathbf{X} , $\|\mathbf{b}\|_2^2 = \sum_{i=1}^p b_i^2$ is the L_2 norm penalty on \mathbf{b} , and
 123 $\lambda > 0$ is the tuning parameter which is associated with the degree of linear
 124 shrinkage. We have the ordinary least squares when $\lambda = 0$. The larger value
 125 of λ leads to the greater amount of shrinkage. However, the ridge \hat{b} 's cannot
 126 be zeros no matter how large the value of λ is set. The value of λ is dependent
 127 on the data, it can be **optimally determined using cross-validation**.

128 *LASSO regression*

LASSO (least absolute shrinkage and selection operator) regression method is widely used in **variable selection** and in the domain with massive dataset [27, 26]. The LASSO performs less sufficient when the features are highly correlated. The method tends to choose a subset of the features, it shrinks some coefficients and sets coefficients of other features to zero. The optimization problem for the LASSO regression estimation with L_1 norm penalty is written as follow:

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \{ \|\mathbf{y} - \mathbf{X}\mathbf{b}\|_2^2 + \lambda \|\mathbf{b}\|_1 \},$$

129 where $\|\mathbf{b}\|_1 = \sum_{i=1}^p |b_i|$ is the L_1 norm, λ is the tuning parameter. L_1 norm
 130 makes LASSO regularize the least squares fit and shrinks some components
 131 to zeros. The suitable value for λ , which is dependent on data, is optimally
 132 selected by cross-validation.

133 *Elastic net regression*

The elastic net regression method is an extension of LASSO that is robust to extreme correlations among the features [28, 29]. The elastic net simultaneously does automatic variable selection and continuous shrinkage, the groups of correlated variables can also be selected. The elastic net uses both L_1 (LASSO) and L_2 (ridge) penalty, the optimization problem is formulated as follow:

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \{ \|\mathbf{y} - \mathbf{X}\mathbf{b}\|_2^2 + \lambda_1 \|\mathbf{b}\|_1 + \lambda_2 \|\mathbf{b}\|_2^2 \}.$$

Let $\alpha = \lambda_2/(\lambda_1 + \lambda_2)$, then the problem is equivalent to solving

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} \|\mathbf{y} - \mathbf{X}\mathbf{b}\|_2^2, \text{ subject to } (1 - \alpha) \|\mathbf{b}\|_1 + \alpha \|\mathbf{b}\|_2^2 \leq t \text{ for some } t.$$

134 The elastic net penalty $(1 - \alpha) \|\mathbf{b}\|_1 + \alpha \|\mathbf{b}\|_2^2 \leq t$ is a convex combination
 135 of the lasso and ridge penalty. The elastic net is a simple ridge regression
 136 when $\alpha = 1$ and a LASSO regression when $\alpha = 0$. The tuning parameter t is
 137 determined with cross-validation for a given α . The L_1 part does automatic
 138 variable selection, while the L_2 part encourages grouped selection [26].

139 *Model validation and performance*

140 With our dataset, we compare the performance of each linear regression
 141 method on fitting the milk production with the model (1). In this experiment,
 142 we fit the linear model using a publicly available R package glmnet [29]. The
 143 values of the tuning parameter are optimized by 10-fold cross-validation and
 144 $\alpha = 0.5$ in the case of the elastic net regression method. The coefficients of
 145 the interesting features fitted by these methods are illustrated in Figure 2.
 146 The coefficient linked to variable starch (kg) is large in all three methods.
 147 The results are reasonable according to the previous studies [30, 13], the
 148 production responded positively to an increment in starch concentration. As
 149 expected, the ridge method keeps all the features, while LASSO and **elastic**
 150 **net shrunk the coefficients of consumption of PDIE (kg) and crude fiber**
 151 **(kg) to zeros. This is due to the correlations between PDIE, crude fiber,**
 152 **Net energy, Starch are high (greater than 0.89).** Table 4 shows the statistical
 153 results of fitting the lactation production with linear regression methods. The
 154 elastic net gives slightly better result, in general, the performance of these
 155 methods are quite similar. In the next part, we will analyze the performance
 156 of the linear model in forecasting the milk production. The comparison with
 157 other machine learning methods will be executed as well.

Statistics	Ridge	LASSO	Elastic net
RMSE	3.23	3.15	3.12
SSE	10753	10240	10054
R^2	0.86	0.87	0.87

Table 4: Statistical values of linear fitting model using Ridge, LASSO and Elastic net. Root Mean Square Error (RMSE), Sum of Squared Errors (SSE), R^2 .

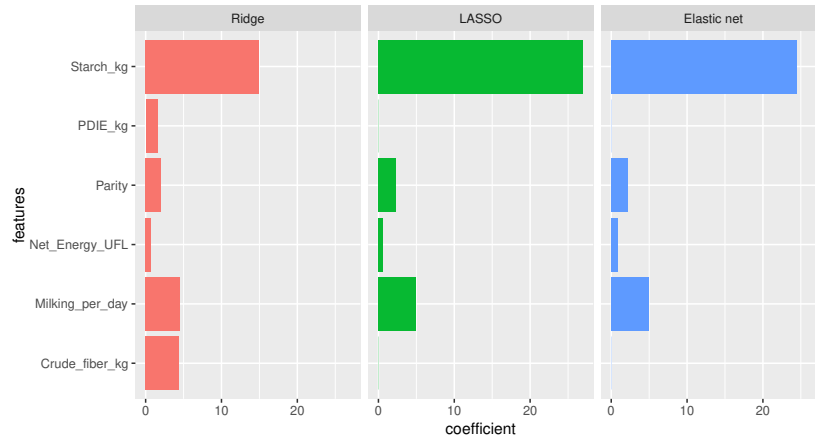


Figure 2: The coefficient of each features estimated by ridge, LASSO, elastic net ($\alpha = 0.5$) regression.

158 3.2. Machine learning algorithms

159 **On forecasting milk production**, in this study, we investigate three ma-
 160 chine learning algorithms: support vector machine regression (SVR), artifi-
 161 cial neural network (ANN), and random forest (RF). These algorithms were
 162 applied in previous studies in the domain of agriculture [31, 32, 33, 34]. **The**
 163 **multiple linear model is also used in the prediction of milk production and**
 164 **compared with these three machine learning algorithms.**

165 *Support vector regression*

The Support Vector Machine is a supervised learning algorithm applied frequently in classification and regression analysis. The Support Vector Machine for function estimation is usually called Support Vector Regression [35]. Suppose we have a training data $\{(x_1, y_1), \dots, (x_n, y_n)\} \in X \times \mathbb{R}$, where X denotes the space of the input features (e.g. $X = \mathbb{R}^d$). In ε -SV regression, the objective is to find a function $f(x)$ that has at most ε deviation from the actual observed data point y_i for all that training data, and is as flat as possible at the same time. In case of a non-linear SVR, the input data are mapped to higher dimensional Hilbert space \mathcal{H} where the regression line can be linearly constructed. For the sake of presentation, a linear regression line

is found by solving the following optimization problem:

$$\begin{aligned} & \underset{\mathbf{w}, \boldsymbol{\xi}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ & \text{subject to} \quad \begin{cases} y_i - \langle \mathbf{w}, x_i \rangle - b \leq \varepsilon + \xi_i, \text{ with } b \in \mathbb{R} \\ \langle \mathbf{w}, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \end{cases} \end{aligned}$$

where \mathbf{w} is the slope of the hyperplane, $\langle \cdot, \cdot \rangle$ denotes the dot product in X . The slack variables ξ_i, ξ_i^* are introduced for the "soft margin" loss function. The constant $C > 0$ determines the trade-off between the flatness of function f and the amount of data points whose deviations are larger than ε are tolerated. Figure 3 graphically interpret a linear SVR. In the non-linear problem, a kernel function k is responsible for computing the dot product in the high dimensional space. In this study, we used the Gaussian or radical basis function (RBF) kernel:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \text{ with } x_i, x_j \in X.$$

The parameters are tuned with the 10-fold cross-validation using the R

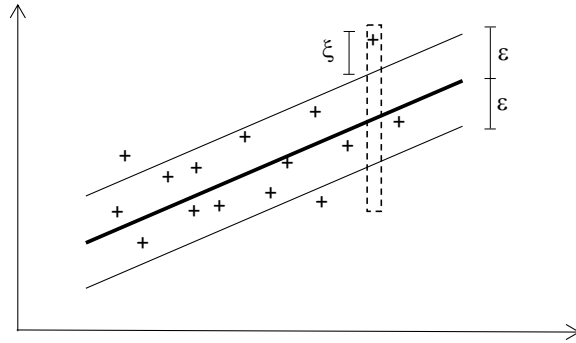


Figure 3: The soft margin loss setting for a linear SVR.

166

167 package 'e1071' [36]. In this dataset, the optimal parameters, in term of
168 smallest mean squared error, are $C = 100, \gamma = 0.01$.

169 *Random forest*

170 Random Forest [37] is an algorithm that learns from multiple decision
171 trees driven on slightly different subsets of data. The random forest algorithm
172 can be applied for both classification and regression. The procedure of the
173 algorithm consists of three stages [38]. The first stage is to create n_{tree}
174 bootstrap samples from the data. Particularly, each sample (bag) contains
175 N observations which are uniformly selected (with replacement) out of N
176 original observations using bootstrap. Then for each sample, we grow a
177 decision CART (Classification and Regression Tree) [39]. Instead of using
178 all predictors, at each node of each tree, m_{try} of the predictors are randomly
179 selected, and the best split is chosen from those variables. Finally, for the
180 new data, the prediction is obtained by aggregating the predictions of the
181 n_{tree} trees, i.e., the average of all prediction of each tree in case of regression.
182 The advantage of the Random Forest is that it can be easily implemented for
183 the nonlinear cases. The R package 'randomForest' ported by Liaw et al. [38]
184 is used in this paper. For our dataset, **by doing three repetitions of 10-fold**
185 **cross-validation, the parameters $n_{tree} = 2000$ and $m_{try} = 4$ are selected.**

186 *Artificial neural network*

187 As the name suggested, this is a connectionist system that is inspired
188 by biological neural networks. It is also commonly known as the multilayer
189 perceptron (MLP). A standard neural network consists of many connected
190 nodes called neural, constructing the input, hidden and output layers. Each
191 neuron produces a sequence of real-value activation. The input values are
192 multiplied by the synaptic weights, which present the strength of the con-
193 nection. The sum of these products is fed to each neuron within the hidden
194 layer via a typically non-linear real-valued activation function such as tanh
195 or logistic [40, 41]. In the case of a single hidden layer, the values are then
196 fed into the output layer neural via the activation function, and predict the
197 output value for each instance. Figure 4 depicts the fully connected artificial
198 neural network. During the training process, MLPs employ backpropagation
199 techniques to minimize the sum of squared errors [42].

200 In this paper, we investigate the fully connected feed-forward neural net-
201 work with one hidden layer; the inputs are parity, DIM, ..., NE; and the
202 output is the milk yield. The R package 'neuralnet' [43] is used to imple-
203 ment the data in our study. **To avoid overfitting the training data, we have**

204 tested few configurations³, and have selected the best by cross-validation.
 205 The optimum network consisted of 4 neurons in the hidden layer is used [9].
 206 The resilient back-propagation with weight backtracking is applied to train
 207 the data. The logistic function in (2) is carried out as the activation function:

$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}. \quad (2)$$

208

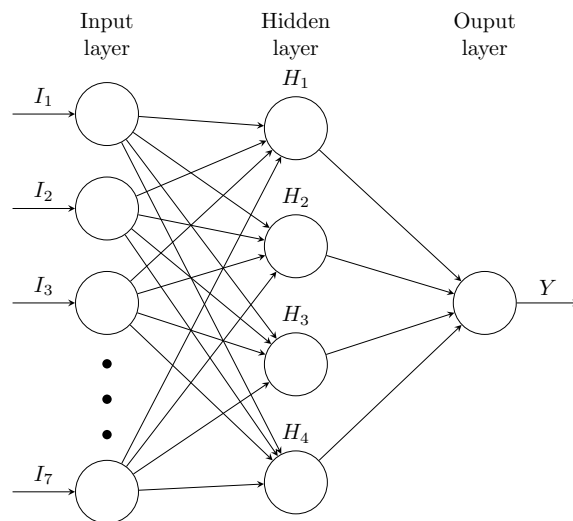


Figure 4: Artificial neural network with one hidden layer.

209 4. Prediction performance comparison and discussion

210 In order to evaluate the prediction performance of the multiple linear
 211 regression (MLR) with elastic regression and the machine learning algorithms
 212 on this dataset; for each cow, the training set is the dataset excluding the data
 213 of one individual. The trained model is then used to predict the production
 214 of the excluded dairy cow. Moreover, the autoregressive versions of these
 215 methods are also investigated in this paper. The evaluation criteria chosen
 216 in this study include: Root Mean Squared Error (RMSE), Mean Absolute

³configurations that have been tested: 4, 5, 6, 7 neurons with Logistic, ReLu activation functions

217 Error (MAE) and **Coefficient of Determination (R^2)**. In addition, we also
218 compare the computational cost of each model to each other.

219 The computer used in this study was a MacBook Pro with Intel core i7 2.5
220 GHz and 16 G 1600 MHz DDR3. Table 5 and Figure 5 present the RMSE, the
221 MAE and the R^2 values of the elastic regression, SVR, random forest, neural
222 network forecasts, **respectively**, against dataset of 36 individual cows in case
223 of no autoregression. **There are some R^2 values that are negative. This is
224 due to the over estimation of the prediction. For instance, as demonstrated
225 in Figure 6, the over predictions of milk yield for the cow #16 make greater
226 error than the mean value does. However, the predictions illustrate well the
227 shape of the observations, the correlation is 0.82. The negative R^2 values
228 were set to $R^2 = 0$ in the subsequent analysis.** The maximum and minimum
229 RMSE values are 5.16 and 1.56 for the MLR, 4.61 and 1.44 for the SVR, 5.77
230 and 1.46 for the random forest, 4.75 and 1.46 for the neural network. Table 6
231 shows the average errors of each model for all 36 individual cows. In general,
232 all the machine algorithms mostly outperform the MLR. The random forest
233 and SVR give the most favorable results, and random forest model is more
234 accurate in term of RMSE and MAE. Moreover, in Table 7, the random forest
235 can compute the internal estimates of variable importance (in percentage).
236 Similar to the **results of MLR model**, starch is the most importance variable
237 according to the random forest algorithm.

PLEASE PUT THE TABLE 5 HERE

PLEASE PUT THE FIGURES 5 HERE

PLEASE PUT THE TABLES 6, 7 HERE

238

239 In addition, in our data collection procedure, there are two cows that
240 were having medical issues. In Figure 6, we present the lactation curves of
241 these two individuals: cow #8 was diagnosed lame at week 24-th of lactation,
242 and cow #9 was diagnosed mastitis at Jun 2016 and August 2016. We can
243 also observe that the production changed at these points, and the predictions
244 become less accurate around these points. Due to the health condition, the
245 amount of food consumption may vary, which leads to the variation in the
246 prediction. This observation is interesting in future studies in detecting the
247 potential health issue of each individual.

PLEASE PUT THE FIGURES 6, 7 HERE

As shown in Table 8, the MLR has the least training time (in seconds) due to its simplicity, while the neural network model has the most expensive computing. **The SVR has a substantial better computational time than the random forest. It also gives better result than the MLR.** Therefore, in term of both accuracy and computational cost, the SVR gives the most sufficient result.

PLEASE PUT THE TABLE 8 HERE

249 A nonlinear autoregressive exogenous (NARX) model has been applied
 250 to milk production forecasting at herd level in the study by Murphy et al.
 251 [9]. In that study, the training data consists of daily herd milk yield, days in
 252 milk and number of cows milked, and the NARX was shown to be the most
 253 effective milk-production model. In our study, the autoregressive version of
 254 the aforementioned models is also considered. The autoregressive models
 255 applied in our experiment have an order of one. In particular, the record in
 256 the previous week is added into the prediction variables:

$$y_t = F(y_{t-1}, u_1, u_2, \dots, u_p) + \varepsilon_t,$$

257 where y_t is the average milk production record **on week t** , and $\{u_1, u_2, \dots, u_p\}$
 258 are the other prediction variables. Table 9 and Figure 8 present the errors
 259 of the autoregressive version of all four forecasting models against dataset
 260 of 36 individual cows. In all cases, the autoregressive approach significantly
 261 improves the accuracy of all prediction models. For example, considering
 262 individual cow ID #7, the RMSEs of four models without autoregression
 263 are 2.44, 2.22, 2.81 and 2.67, respectively; with autoregression, the errors
 264 decreased to 1.88, 1.89, 2.35 and 1.80, respectively. However, considering the
 265 cow number 35, we get more error with the autoregressive models, this can be
 266 caused by the status of that individual (e.g. health problem). Therefore, milk
 267 yield forecasting could be applied in monitoring health conditions [14]. In
 268 average, Table 10 show a **substantial** improvement in accuracy compared to
 269 the model without autoregression, **the R^2 values of the regression are mostly**
 270 **high**. Moreover, as shown in Table 11, the internal estimates of variable
 271 importance computed by random forest show that the information in the
 272 past is essentially important (62.78%), starch is still an important variable
 273 (14.81%) compared to the rest.

PLEASE PUT THE TABLE 9 HERE

PLEASE PUT THE FIGURE 8 HERE

PLEASE PUT THE TABLES 10, 11 HERE

Table 12 presents the average training time for the autoregressive model, the random forest and neural network still consume more computing power than the MLR and SVR. The SVR is yet the best compromise between accuracy and computational cost. In practice, with a portable application, the dairy farmers can improve and update the database in realtime, and train the model with the local dataset. Therefore, **it is potentially suitable for industrial applications.**

PLEASE PUT THE TABLE 12 HERE

274

275 5. Concluding remarks

276 This is a study on a small scale (36 milking cows) in Brittany, France.
277 The correlation between fat and protein content and milk yield with the col-
278 lected data has indicated the decrease of the fat and protein content as milk
279 yield increases to 20 (kg/day). On this dataset, **the analysis of the chemical**
280 **composition of nutrition has shown the significant weight of nutrition supply**
281 **through the diet on the milk production level of dairy cattle, which is more**
282 **important than milk per day and parity.**

283 Moreover, we compare the performance of the linear regression models
284 and machine learning models on forecasting milk production at the individ-
285 ual level. **For each model, we investigate both versions: autoregressive and**
286 **non-autoregressive approaches. With this dataset, the autoregressive mod-**
287 **els, which consider the previous observation,** are shown to be significantly
288 better than the non-autoregressive approaches. When the past is consid-
289 ered, the information from the previous observation considerably improves
290 the prediction accuracy.

291 **Among the different methods, the random forest gives the best perfor-**
292 **mance on 15 individuals, the support vector machine gives prediction with**
293 **the smallest errors on 13 dairy cows. The linear and neural network models**
294 **show the best results on 5 and 3 individuals, respectively. However, the com-**
295 **putational times of SVR are significantly less than random forest. Therefore,**
296 **the support vector regression is the most efficient method for predicting milk**
297 **production** among the other models in terms of both prediction accuracy and

298 computational cost. The result indicates the possibility of practical appli-
299 cation on a small scale farm with a small number of dairy cows. However,
300 the autoregressive models require the previous observation, then the non-
301 autoregressive approaches are more practical when past observations are not
302 available, or a far prediction is considered. Further research on other kinds
303 of dairy cows with **larger cow population sizes over longer time periods** is re-
304 quired to investigate the potential of using these models in health monitoring
305 on an individual cow level with high accuracy.

306 **Acknowledgments**

307 This research activity have been financed by Conseil regional Bretagne
308 and FEDER Bretagne within the project NUTGEN of the Université de
309 Bretagne Sud.

310 **References**

- 311 [1] P. Wood, Algebraic model of the lactation curve in cattle, *Nature* 216
312 (1967) 164–165.
- 313 [2] T. Ali, L. Schaeffer, Accounting for covariances among test day milk
314 yields in dairy cows, *Canadian Journal of Animal Science* 67 (1987)
315 637–644.
- 316 [3] J. Wilmink, Adjustment of lactation yield for age at calving in relation
317 to level of production, *Livestock Production Science* 16 (1987) 321–334.
- 318 [4] L. Schaeffer, Application of random regression models in animal breed-
319 ing, *Livestock Production Science* 86 (2004) 35–45.
- 320 [5] A. Silvestre, A. Martins, V. Santos, M. Ginja, J. Colaço, Lactation
321 curves for milk, fat and protein in dairy cows: A full approach, *Livestock*
322 *Science* 122 (2009) 308–313.
- 323 [6] S. Adediran, D. Ratkowsky, D. Donaghy, A. Malau-Aduli, Comparative
324 evaluation of a new lactation curve model for pasture-based holstein-
325 friesian dairy cows, *Journal of dairy science* 95 (2012) 5344–5356.

- 326 [7] M. Mellado, J. Flores, A. De Santiago, F. Veliz, U. Macías-Cruz,
327 L. Avendaño-Reyes, J. García, Extended lactation in high-yielding hol-
328 stein cows: Characterization of milk yield and risk factors for lactations
329 > 450 days, *Livestock Science* 189 (2016) 50–55.
- 330 [8] J. O. Lehmann, L. Mogensen, T. Kristensen, Extended lactations in
331 dairy production: Economic, productivity and climatic impact at herd,
332 farm and sector level, *Livestock science* 220 (2019) 100–110.
- 333 [9] M. Murphy, M. O’Mahony, L. Shalloo, P. French, J. Upton, Comparison
334 of modelling techniques for milk-production forecasting, *Journal of dairy
335 science* 97 (2014) 3352–3363.
- 336 [10] F. Zhang, M. D. Murphy, L. Shalloo, E. Ruelle, J. Upton, An automatic
337 model configuration and optimization system for milk production fore-
338 casting, *Computers and Electronics in Agriculture* 128 (2016) 100–111.
- 339 [11] S. Nickerson, Milk production: Factors affecting milk composition, in:
340 *Milk quality*, Springer, 1995, pp. 3–24.
- 341 [12] I. Harder, E. Stamer, W. Junge, G. Thaller, Lactation curves and model
342 evaluation for feed intake and energy balance in dairy cows, *Journal of
343 Dairy Science* (2019).
- 344 [13] J. Boerman, S. Potts, M. VandeHaar, M. Allen, A. Lock, Milk pro-
345 duction responses to a change in dietary starch concentration vary by
346 production level in dairy cattle, *Journal of dairy science* 98 (2015) 4698–
347 4706.
- 348 [14] F. Andersen, O. Østerås, O. Reksen, Y. T. Gröhn, Mastitis and the
349 shape of the lactation curve in norwegian dairy cows, *Journal of dairy
350 research* 78 (2011) 23–31.
- 351 [15] D. B. Jensen, M. van der Voort, H. Hogeveen, Dynamic forecasting of
352 individual cow milk yield in automatic milking systems, *Journal of dairy
353 science* 101 (2018) 10428–10439.
- 354 [16] F. Zhang, J. Upton, L. Shalloo, M. Murphy, Effect of parity weight-
355 ing on milk production forecast models, *Computers and Electronics in
356 Agriculture* 157 (2019) 589–603.

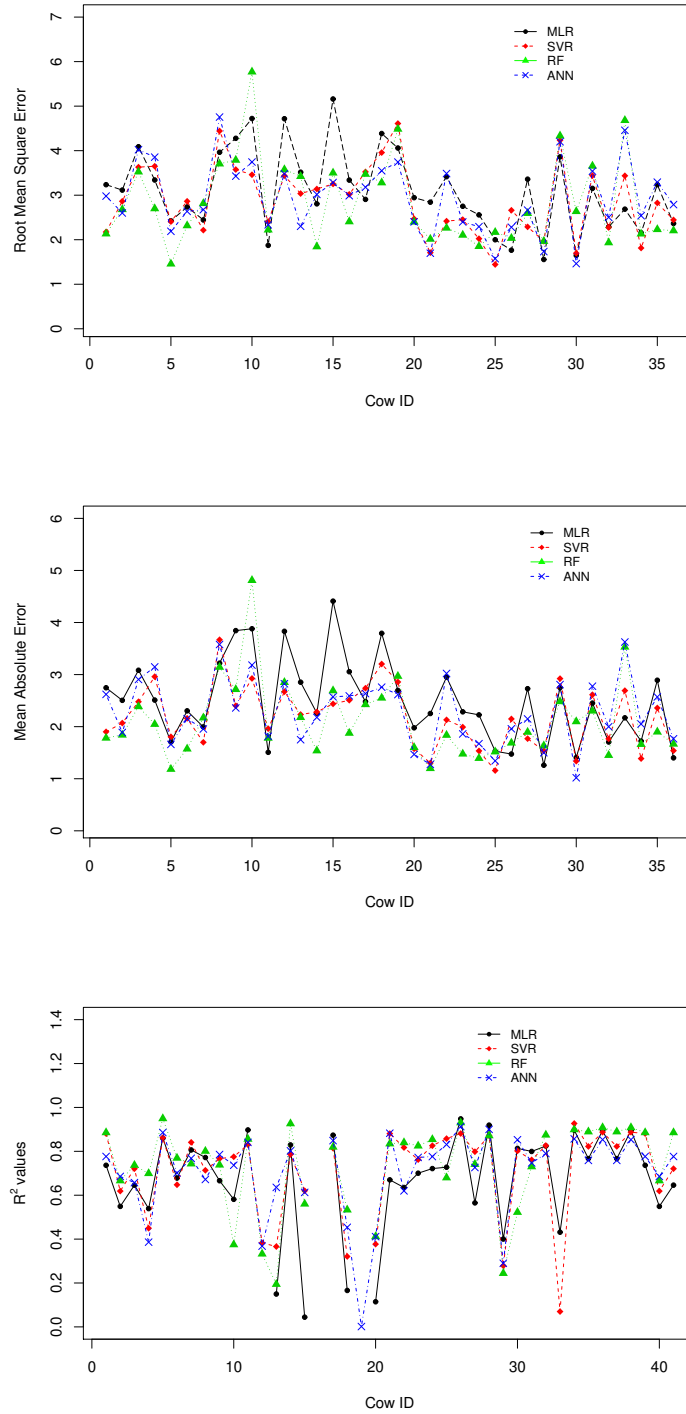
- 357 [17] J. Van Bebber, N. Reinsch, W. Junge, E. Kalm, Monitoring daily milk
358 yields with a recursive test day repeatability model (kalman filter), *Journal of dairy science* 82 (1999) 2421–2429.
359
- 360 [18] M. Vermorel, Energy: the feed unit systems, in: *Ruminant nutrition: recommended allowances and feed tables*, INRA Publications, Paris, 1989, p. 28.
361
362
- 363 [19] S. A. Kadi, F. Djellal, M. Berchiche, Caractérisation de la conduite ali-
364 mentaire des vaches laitières dans la région de tizi-ouzou, algérie, *Livestock Research for rural development* 19 (2007).
365
- 366 [20] G. Pulina, A. Nudda, G. Battacone, A. Cannas, Effects of nutrition on
367 the contents of fat, protein, somatic cells, aromatic compounds, and un-
368 desirable substances in sheep milk, *Animal Feed Science and Technology*
369 131 (2006) 255–291.
- 370 [21] R. Emery, Milk fat depression and the influence of diet on milk compo-
371 sition., *The Veterinary Clinics of North America. Food Animal Practice*
372 4 (1988) 289–305.
- 373 [22] U. Meyer, M. Everinghoff, D. Gädeken, G. Flachowsky, Investigations
374 on the water intake of lactating dairy cows, *Livestock production science*
375 90 (2004) 117–121.
- 376 [23] G. Pulina, N. Macciotta, A. Nudda, Milk composition and feeding in
377 the italian dairy sheep, *Italian Journal of Animal Science* 4 (2005) 5–14.
- 378 [24] V. Olori, S. Brotherstone, W. Hill, B. McGuirk, Fit of standard models
379 of the lactation curve to weekly records of milk production of cows in a
380 single herd, *Livestock Production Science* 58 (1999) 55–63.
- 381 [25] A. E. Hoerl, R. W. Kennard, Ridge regression: Biased estimation for
382 nonorthogonal problems, *Technometrics* 12 (1970) 55–67.
- 383 [26] J. O. Ogutu, T. Schulz-Streeck, H.-P. Piepho, Genomic selection using
384 regularized linear regression models: ridge regression, lasso, elastic net
385 and their extensions, *BMC Proceedings* 6 (2012) S10.
- 386 [27] R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal*
387 *of the Royal Statistical Society: Series B (Methodological)* 58 (1996)
388 267–288.

- 389 [28] H. Zou, T. Hastie, Regularization and variable selection via the elastic
390 net, *Journal of the royal statistical society: series B (statistical method-*
391 *ology)* 67 (2005) 301–320.
- 392 [29] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for general-
393 ized linear models via coordinate descent, *Journal of statistical software*
394 33 (2010) 1.
- 395 [30] A. Cabrita, R. Bessa, S. Alves, R. Dewhurst, A. Fonseca, Effects of
396 dietary protein and starch on intake, milk production, and milk fatty
397 acid profiles of dairy cows fed corn silage-based diets, *Journal of Dairy*
398 *Science* 90 (2007) 1429–1439.
- 399 [31] C. Kamphuis, H. Mollenhorst, A. Feelders, D. Pietersma, H. Hogeveen,
400 Decision-tree induction to detect clinical mastitis with automatic milk-
401 ing, *Computers and Electronics in Agriculture* 70 (2010) 60–68.
- 402 [32] K. Saruta, Y. Hirai, K. Tanaka, E. Inoue, T. Okayasu, M. Mitsuoka,
403 Predictive models for yield and protein content of brown rice using sup-
404 port vector machine, *Computers and electronics in agriculture* 99 (2013)
405 93–100.
- 406 [33] B. Barrett, I. Nitze, S. Green, F. Cawkwell, Assessment of multi-
407 temporal, multi-sensor radar and ancillary spatial data for grasslands
408 monitoring in ireland using machine learning approaches, *Remote Sens-*
409 *ing of Environment* 152 (2014) 109–124.
- 410 [34] P. Shine, M. D. Murphy, J. Upton, T. Scully, Machine-learning algo-
411 rithms for predicting on-farm direct water and electricity consumption
412 on pasture based dairy farms, *Computers and electronics in agriculture*
413 150 (2018) 74–87.
- 414 [35] A. J. Smola, B. Schölkopf, A tutorial on support vector regression,
415 *Statistics and computing* 14 (2004) 199–222.
- 416 [36] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, C.-C.
417 Chang, C.-C. Lin, M. D. Meyer, Package ‘e1071’, *The R Journal* (2019).
- 418 [37] L. Breiman, Random forests, *Machine learning* 45 (2001) 5–32.

- 419 [38] A. Liaw, M. Wiener, et al., Classification and regression by randomfor-
420 est, *R news* 2 (2002) 18–22.
- 421 [39] L. Breiman, *Classification and regression trees*, Routledge, 2017.
- 422 [40] P. Bickel, P. Diggle, S. Fienberg, U. Gather, I. Olkin, S. Zeger, *Springer*
423 *Series in Statistics*, Springer, 2009.
- 424 [41] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural*
425 *networks* 61 (2015) 85–117.
- 426 [42] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hub-
427 bard, L. D. Jackel, Backpropagation applied to handwritten zip code
428 recognition, *Neural computation* 1 (1989) 541–551.
- 429 [43] F. Günther, S. Fritsch, neuralnet: Training of neural networks, *The R*
430 *journal* 2 (2010) 30–38.

Cow ID	MLR			SVR			Random forest			Neural network		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
1	3.24	2.75	0.74	2.17	1.90	0.88	2.13	1.78	0.89	2.98	2.62	0.78
2	3.11	2.51	0.55	2.86	2.07	0.62	2.68	1.84	0.67	2.6	1.89	0.69
3	4.09	3.08	0.65	3.63	2.48	0.72	3.52	2.39	0.74	4.02	2.91	0.66
4	3.34	2.51	0.54	3.65	2.96	0.45	2.70	2.05	0.70	3.85	3.15	0.39
5	2.42	1.71	0.86	2.41	1.81	0.86	1.46	1.18	0.95	2.19	1.66	0.89
6	2.74	2.30	0.68	2.86	2.16	0.65	2.32	1.57	0.77	2.64	2.15	0.70
7	2.44	1.99	0.81	2.22	1.70	0.84	2.81	2.17	0.74	2.67	1.96	0.77
8	3.96	3.22	0.77	4.44	3.67	0.71	3.70	3.14	0.80	4.75	3.58	0.67
9	4.28	3.84	0.67	3.58	2.40	0.77	3.79	2.72	0.74	3.43	2.36	0.79
10	4.72	3.88	0.58	3.46	2.93	0.78	5.77	4.81	0.37	3.75	3.18	0.74
11	1.87	1.51	0.90	2.41	1.96	0.83	2.22	1.78	0.86	2.33	1.82	0.84
12	4.72	3.83	0	3.44	2.67	0.38	3.58	2.85	0.33	3.48	2.82	0.37
13	3.52	2.85	0.15	3.04	2.23	0.37	3.42	2.18	0.19	2.3	1.75	0.64
14	2.81	2.26	0.83	3.14	2.28	0.79	1.84	1.54	0.93	3.02	2.18	0.8
15	5.16	4.41	0.04	3.25	2.44	0.62	3.5	2.69	0.56	3.28	2.57	0.61
16	3.34	3.06	0	3.02	2.51	0	2.41	1.88	0	2.99	2.59	0
17	2.91	2.48	0.87	3.52	2.74	0.81	3.47	2.43	0.82	3.18	2.63	0.85
18	4.38	3.79	0.17	3.96	3.2	0.32	3.28	2.55	0.53	3.55	2.76	0.45
19	4.06	2.70	0	4.61	2.86	0	4.49	2.97	0	3.74	2.62	0.001
20	2.94	1.98	0.11	2.47	1.58	0.38	2.40	1.60	0.41	2.40	1.47	0.41
21	2.84	2.25	0.67	1.71	1.30	0.88	2.01	1.20	0.83	1.70	1.27	0.88
22	3.42	2.95	0.64	2.42	2.13	0.82	2.26	1.84	0.84	3.49	3.02	0.62
23	2.75	2.28	0.70	2.45	1.99	0.76	2.10	1.48	0.82	2.40	1.86	0.77
24	2.56	2.23	0.72	2.02	1.53	0.83	1.85	1.39	0.85	2.29	1.67	0.78
25	2.00	1.53	0.73	1.44	1.16	0.86	2.17	1.52	0.68	1.57	1.34	0.83
26	1.76	1.47	0.95	2.66	2.15	0.88	2.03	1.69	0.93	2.28	1.96	0.91
27	3.36	2.73	0.57	2.29	1.77	0.80	2.59	1.90	0.74	2.67	2.15	0.73
28	1.56	1.26	0.92	1.97	1.54	0.87	1.96	1.63	0.87	1.73	1.50	0.90
29	3.86	2.75	0.40	4.23	2.92	0.28	4.33	2.48	0.24	4.20	2.81	0.29
30	1.65	1.41	0.81	1.70	1.34	0.80	2.64	2.10	0.52	1.46	1.02	0.85
31	3.15	2.45	0.80	3.44	2.61	0.76	3.66	2.30	0.73	3.54	2.78	0.75
32	2.29	1.71	0.82	2.28	1.77	0.83	1.93	1.45	0.87	2.50	2.00	0.79
33	2.69	2.17	0.43	3.44	2.69	0.07	4.68	3.53	0	4.46	3.63	0
34	2.16	1.72	0.90	1.81	1.39	0.93	2.13	1.66	0.90	2.54	2.05	0.86
35	3.24	2.89	0.77	2.83	2.36	0.82	2.23	1.90	0.89	3.29	2.57	0.76
36	2.36	1.40	0.89	2.45	1.54	0.89	2.20	1.67	0.91	2.79	1.77	0.85

Table 5: The forecast error of four models for 36 individual cows.



23
 Figure 5: Root Mean Squared Error (above), Mean Absolute Error (middle) and R^2 values (below) of four model forecasts for 36 individual cows.

	Elastic regression	SVR	Random forest	Neural Network
RMSE	3.103	2.868	2.842	2.947
MAE	2.496	2.187	2.107	2.279
R^2	0.664	0.712	0.734	0.704

Table 6: Average error of each model for all 36 individual cows.

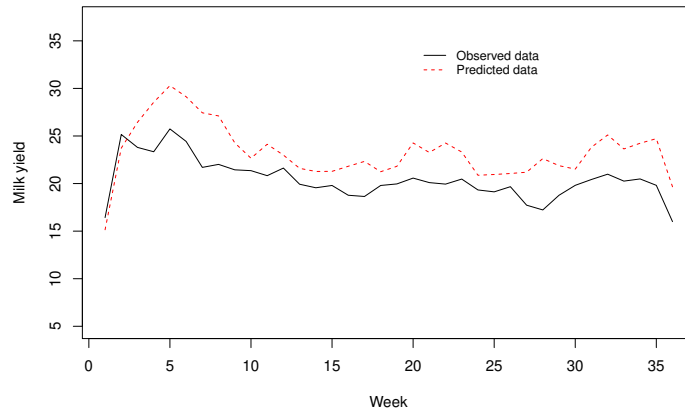


Figure 6: The observations and predictions of milk production of cow number 16 using MLR, R^2 value is -1.46.

Parity	DIM	MPD	Starch	Crude fiber	PDIE	NE
11.21	8.96	11.35	34.87	15.44	6.58	15.16

Table 7: Average of variable importance estimated by random forest (in %).

	Elastic regression	SVR	Random forest	Neural Network
mean	0.077	0.157	6.771	7.357
SD	0.005	0.007	0.175	4.754

Table 8: Average training time (in seconds) and its standard deviation for 36 experiments.

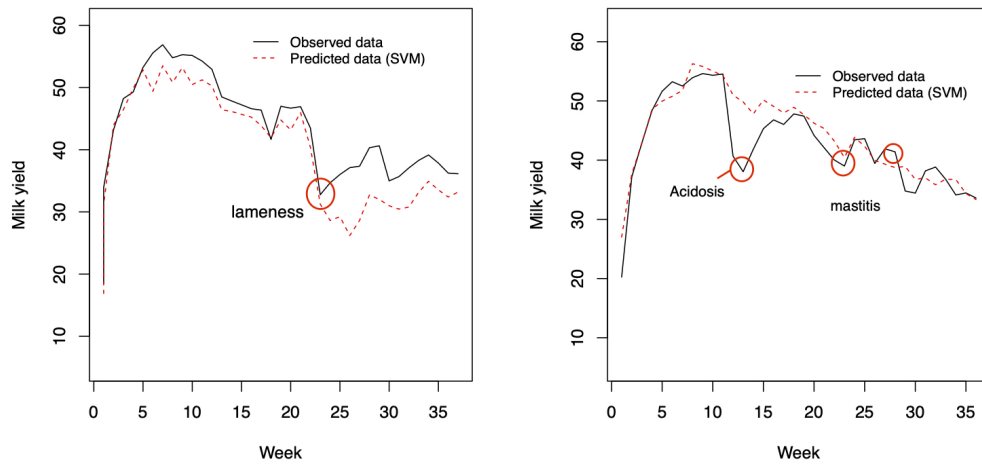


Figure 7: Two individual cows that had medical issues during the experiment, one had lameness (left), while the other had mastitis (right).

Cow ID	MLR			SVR			Random forest			Neural network		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
1	1.93	1.72	0.91	1.36	1.09	0.95	1.37	1.07	0.95	1.34	1.14	0.95
2	1.81	1.55	0.85	1.72	1.27	0.86	2.14	1.57	0.79	1.81	1.49	0.85
3	3.79	2.61	0.70	4.20	2.80	0.63	4.26	2.95	0.62	3.51	2.28	0.74
4	2.69	2.23	0.70	2.75	2.38	0.69	2.17	1.45	0.81	2.86	2.41	0.66
5	1.72	1.30	0.93	1.81	1.37	0.92	1.47	1.19	0.95	1.84	1.38	0.92
6	2.15	1.59	0.80	2.20	1.30	0.79	1.51	1.13	0.90	2.61	1.53	0.71
7	1.88	1.56	0.89	1.89	1.57	0.88	2.35	1.85	0.82	1.80	1.49	0.89
8	2.48	1.74	0.91	3.61	3.05	0.81	3.06	2.48	0.86	2.55	1.83	0.91
9	3.12	2.26	0.82	2.71	1.74	0.87	3.15	2.24	0.82	2.93	2.08	0.84
10	3.17	2.4	0.81	2.91	2.47	0.84	3.62	2.76	0.75	4.12	3.76	0.68
11	1.60	1.28	0.93	1.78	1.27	0.91	1.32	1.02	0.95	1.90	1.44	0.89
12	2.76	2.07	0.60	2.26	1.51	0.73	2.46	1.90	0.68	2.47	1.70	0.68
13	2.62	2.08	0.53	2.44	2.01	0.59	2.74	2.19	0.49	2.61	2.16	0.53
14	1.91	1.62	0.92	2.63	2.12	0.85	2.10	1.71	0.90	2.36	2.00	0.88
15	3.29	2.77	0.61	2.68	2.00	0.74	2.71	2.06	0.74	2.77	2.31	0.72
16	2.08	1.78	0.04	1.67	1.29	0.38	1.85	1.39	0.24	1.63	1.35	0.42
17	1.77	1.44	0.95	2.07	1.61	0.94	2.03	1.54	0.94	1.97	1.65	0.94
18	2.50	1.95	0.73	2.08	1.43	0.81	2.36	1.64	0.76	2.54	1.89	0.72
19	2.60	1.82	0.52	3.18	2.23	0.28	2.82	2.06	0.43	3.10	2.08	0.32
20	1.66	1.30	0.72	1.29	1.06	0.83	1.46	1.16	0.78	1.67	1.32	0.71
21	2.31	1.53	0.78	1.89	1.13	0.85	1.58	1.02	0.90	1.75	1.21	0.88
22	2.10	1.71	0.86	1.55	1.28	0.92	1.58	1.33	0.92	1.77	1.48	0.90
23	1.98	1.44	0.84	1.85	1.23	0.86	1.84	1.24	0.87	1.94	1.43	0.85
24	1.58	1.32	0.89	1.56	1.16	0.90	1.27	1.05	0.93	1.69	1.29	0.88
25	1.79	1.45	0.78	1.63	1.28	0.82	1.81	1.28	0.78	1.66	1.37	0.81
26	2.57	1.89	0.89	2.74	2.15	0.87	2.24	1.88	0.92	2.97	2.25	0.85
27	2.03	1.68	0.84	1.24	0.95	0.94	1.10	0.79	0.95	1.43	1.19	0.92
28	1.97	1.46	0.87	2.08	1.43	0.86	2.13	1.37	0.85	1.85	1.30	0.89
29	2.43	1.79	0.76	2.66	1.62	0.72	2.82	1.50	0.68	3.42	1.91	0.53
30	1.46	1.23	0.85	1.42	1.10	0.86	1.61	1.11	0.82	1.46	1.22	0.85
31	2.98	2.33	0.82	2.77	2.14	0.84	2.21	1.72	0.90	2.90	2.20	0.83
32	1.77	1.39	0.90	1.81	1.36	0.89	1.30	1.04	0.94	1.88	1.53	0.88
33	1.59	1.28	0.80	1.85	1.56	0.73	1.90	1.42	0.71	1.67	1.49	0.78
34	1.55	1.14	0.95	1.45	1.00	0.95	1.77	1.29	0.93	1.77	1.28	0.93
35	4.48	2.42	0.55	4.38	2.41	0.57	4.17	2.35	0.61	4.61	2.39	0.53
36	2.40	1.69	0.89	2.22	1.72	0.91	2.03	1.52	0.92	2.61	1.88	0.87

Table 9: The forecast error of four autoregressive models for 36 individual cows.

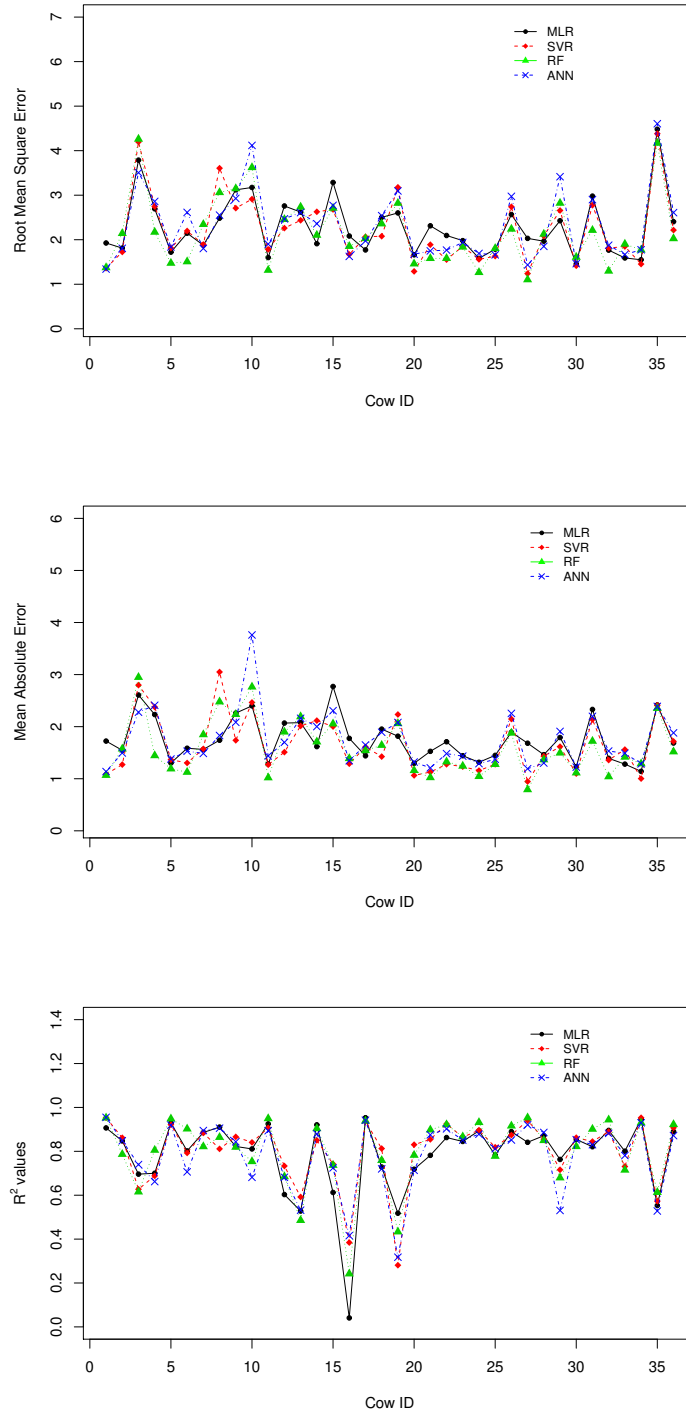


Figure 8: Root Mean Squared Error (above), Mean Absolute Error (middle) and R^2 values (below) of four autoregressive models of order 1 forecasts for 36 individual cows.

	Elastic regression	SVR	Random forest	Neural Network
RMSE	2.292	2.231	2.175	2.327
MAE	1.745	1.641	1.590	1.742
R^2	0.782	0.801	0.801	0.782

Table 10: Average error of each autoregressive model for all 36 individual cows.

Parity	DIM	MPD	Starch	Crude fiber	PDIE	NE	y_{t-1}
2.32	4.28	5.40	14.81	5.11	2.62	6.29	62.78

Table 11: Average of variable importance estimated by random forest (in %).

	Elastic regression	SVR	Random forest	Neural Network
mean	0.083	0.182	7.240	6.862
SD	0.009	0.007	0.152	2.919

Table 12: Average training time (in seconds) and its standard deviation for 36 experiments.