



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. 2019. hal-02358044v1

HAL Id: hal-02358044

<https://hal.science/hal-02358044v1>

Preprint submitted on 11 Nov 2019 (v1), 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 relation between fat/protein content and milk yield was conducted. The fat and protein production increase at a slower rate as milk yield increases. The importance of chemical composition and previous observation on milk production forecasting is investigated using both linear models and machine learning approaches (SVM, random forest, neural network). This study evaluates the prediction accuracy of linear model and machine learning based models in both non-autoregressive and autoregressive cases at the individual level. The autoregressive models have proven to have a better performance than the non-autoregressive approaches. Moreover, the computational cost of each approach is presented in the paper. The support vector machine is shown to be the best compromise between accuracy and computational cost.

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

1. Introduction

2 Milk production forecasting of daily cow is an essential factor which is
3 useful for the daily farmers in management as well as health monitoring.

*Corresponding author

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

4 In literature, many parametric models have been developed to model the
5 lactation curve at herd and individual level [1, 2, 3, 4, 5, 6]. Or the studies
6 on extended lactation in dairy production [7, 8]. Recently, there are a number
7 of modeling techniques on milk production forecasting that showed to obtain
8 a high accurate prediction with adaptability at the herd level [9, 10, 8]. The
9 nonlinear auto regressive model with exogenous input using artificial neural
10 networks introduced by Murphy et al. [9] shows to be most effective milk-
11 production model.

12 Moreover, understanding the effect of the nutritional diet characteristics
13 on the milk production and the quality of milk is not only helpful in financial
14 planning but also in the production of other daily products, such as yogurt,
15 cheese, butter [11]. The importance of feed intake, diet on dairy cows was
16 investigated in recent years. For example, the feed intake increases slowly at
17 the beginning of lactation [12]; or the effects of dietary starch concentration
18 on yield of milk and milk components was investigate by Boerman et al.
19 [13]. In spite of that, not many researches are on individual cow level, and
20 the forecasting based on the nutrition for the small scale farm. Milk yield
21 forecasting of each individual cow can be beneficial to many applications such
22 as monitoring health conditions and disease detection, i.e. mastitis [14, 15].
23 For instance, a study on effect of parity weighting was conducted by Zhang
24 et al. [16] with the dataset in the south of Ireland; or Van Bebber et al. [17]
25 applied Kalman Filter in monitoring daily milk yields.

26 The subject of this study is to improve the livestock farming, particu-
27 larly the milk production, by monitoring the performance in nutrition sup-
28 plies. The first objective considers the importance of chemical composition
29 of nutrition on the production and production monitoring of dairy cattle in
30 Brittany, France. Secondly, we compare different types of multiple linear
31 regression and machine learning based models for prediction of production
32 of individual cow. The practicability and ability for commercial application
33 are also discussed.

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

40 2. Data description and composition analysis

41 2.1. Data description

42 The empirical data were collected from 36 lactating Holstein-Friesian
43 dairy cows in a research farm in Brittany, France, equipped with a robotic
44 milking system. For a ten months period (from December 2015 to September
45 2016), there are 7691 valid milking records collected. Each milking record
46 contains: daily milk yield (DMY), day in milk (DIM), parity information
47 (first, second, third onward lactation, see Tab. 1), number of milking per
48 day and the collective (corn silage, grass silage, wheat straw, soybean meal)
49 or individual (pelleted feed distributed through an automatic feeder) con-
50 sumption of diet components. Each cow is milked one to four times per
51 day by robotic milking system, the cow can possibly be milked each time it
52 comes to the freestall for food. In this experiment, the amount of given diet
53 are changed every week. In this study, we are interested in the effect of the
54 diet on the milk production forecasting. Particularly, the chemical compo-
55 sition studied in this paper are: starch, crude fiber, net energy (NE) Unité
56 Fourragère Lait (UFL¹) and protein (PDIE²). Therefore, the consumption
57 of different diets was converted to these four chemical composition. Table
58 2 presents the composition of each diet. It should be noted that, in Table
59 2, the consumption of the first eight diets (Corn silage, Grass silage, ..., Ni-
60 trogen supplement) is the same for 36 dairy cows at a specific week. On
61 the other hand, since the last four components (Production feed, ..., Liquid
62 feed) in Table 2 are distributed by robot, which means the consumption of
63 these four components varies according to the milk production level of each
64 individual cow. Therefore, the consumption of each individual may differ at
65 a specific week. In order to have a regular effect of each nutrient on the milk
66 production, we used the weekly data instead of the daily data. That means
67 each data point is the average of seven days observations. The statistical
68 characteristics of the interested variables are presented in Table 3.

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

²**P**rotéines **D**igestibles dans l'**I**ntestin limitantes par l'apport d'**É**nergie: true protein absorbable in the small intestine when rumen fermentable energy (organic matter) is limiting microbial protein synthesis in the rumen [19].

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.

69 2.2. Milk fat and protein composition analysis

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

	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.

76 relation coefficients between milk yield and fat and protein content are -0.04
77 and -0.21 , respectively. In our observed data, the fat and protein content
78 decrease as the milk yield increase, but not significant. As shown in Figures
79 1a, 1c, the fat and protein content visually decrease as milk yield increase
80 to 20 (kg/day). This phenomenon can be explained as at the beginning of
81 the lactation, the milk production increases more rapidly than the ability of
82 consumption of the cow. Moreover, when the dairy cows produce more milk,
83 they consume more, specially water [22], but the nutrition absorption cannot
84 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:

$$y = ax^b$$

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

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

101 3. Modelization

102 3.1. Linear Model

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

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

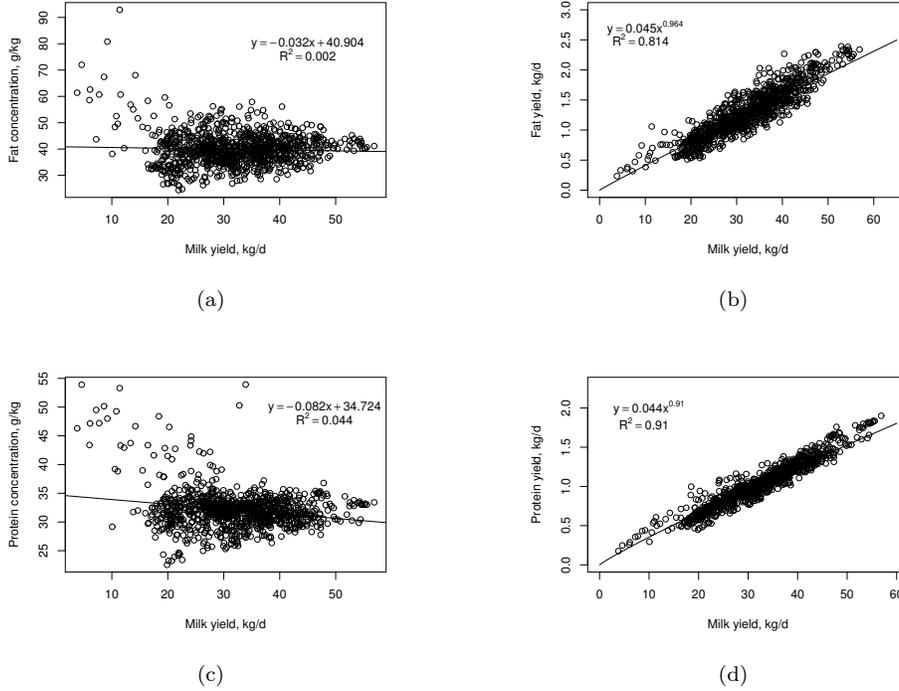


Figure 1: Relationships between milk yield and (a) milk fat yield, (b) milk fat concentration, (c) milk protein yield and (d) milk protein concentration.

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 predictor [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

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

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

114 *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:

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

115 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
 116 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
 117 $\lambda > 0$ is the tuning parameter which is associated with the degree of linear
 118 shrinkage. We have the ordinary least squares when $\lambda = 0$. The larger value
 119 of λ leads to the greater amount of shrinkage. However, the ridge \hat{b} 's cannot
 120 be zeros no matter how large the value of λ is set. The value of λ is dependent
 121 on the data, it can be determined using cross-validation.

122 *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:

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

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

127 *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

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

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

133 *Model validation and performance*

134 With our dataset, we compare the performance of each linear regression
135 method on fitting the milk production with the model (3.1). In this experi-
136 ment, we fit the linear model using a publicly available R package glmnet [29].
137 The values of tuning parameter are optimized by 10-fold cross-validation, and
138 $\alpha = 0.5$ in case of elastic net regression method. The coefficients of the in-
139 terested features fitted by these methods are illustrated in Figure 2. The
140 coefficient linked to variable starch (kg) is large in all three methods. The
141 results are reasonable according to the previous studies [30, 13], the pro-
142 duction responded positively to an increment in starch concentration. As

143 expected, ridge keeps all the features, while LASSO and elastic net shrunk
 144 the coefficients of consumption of PDIE (kg) and crude fiber (kg) to zeros.
 145 Table 4 shows the statistical results of fitting the lactation production with
 146 linear regression methods. The elastic net gives slightly better result, in
 147 general, the performance of these methods are quite similar. In the next
 148 part, we will analyze the performance of the linear model in forecasting the
 149 milk production. The comparison with other machine learning methods is
 150 executed as well.

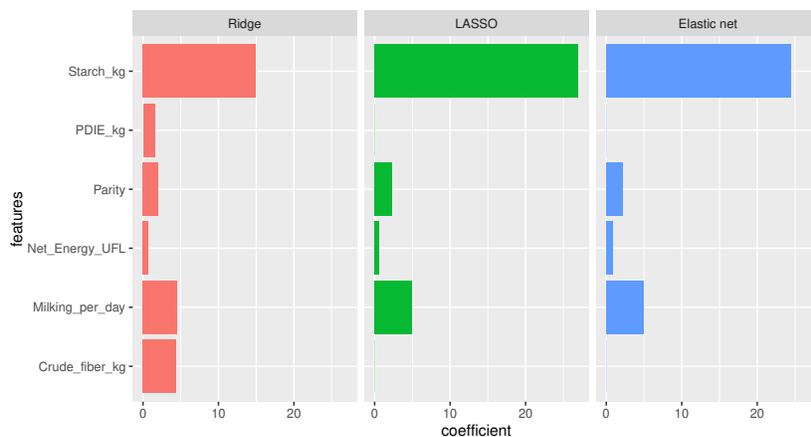


Figure 2: The coefficient of each features estimated by ridge, LASSO, elastic net ($\alpha = 0.5$) regression.

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 .

151 3.2. Machine learning algorithms

152 In this study, we investigate three ML algorithms: support vector machine
 153 regression (SVR), artificial neural network (ANN), and random forest (RF)
 154 in the application of forecasting milk production. These algorithms were
 155 applied in the previous studies in the domain of agriculture [31, 32, 33, 34].

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 (SVR) [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:

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

157 The parameters are tuned with the 10-fold cross-validation using the R
158 package 'e1071' [36]. In this dataset, the optimal parameters, in term of
159 smallest mean squared error, are $C = 100, \gamma = 0.01$.

160 *Random forest*

161 Random Forest [37] is an algorithm that learns from multiple decision
162 trees driven on slightly different subsets of data. The random forests algo-
163 rithm can be applied for both classification and regression. The procedure

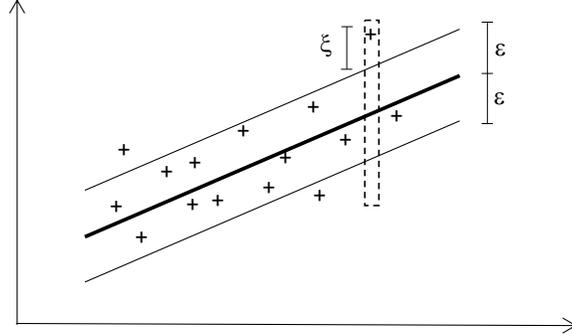


Figure 3: The soft margin loss setting for a linear SVR.

164 of the algorithm consists of three stages [38]. First stage is to create n_{tree}
 165 bootstrap samples from the data. Particularly, each sample (bag) contains
 166 N observations which are uniformly selected (with replacement) out of N
 167 original observations using bootstrap. Then for each sample, we grow a
 168 decision CART (Classification and Regression Tree) [39]. Instead of using all
 169 predictors, at each node of each tree, m_{try} of the predictors are randomly
 170 selected, and the best split is chosen from those variables. Finally, for the
 171 new data, the prediction is obtain by aggregating the predictions of the n_{tree}
 172 trees, i.e., the average of all prediction of each tree in case of regression. The
 173 advantage of the Random Forest is that it can be easily implemented for the
 174 nonlinear cases. The R package 'randomForest' ported by Liaw et al. [38]
 175 is used in this paper. For our dataset, by doing three repetitions of 10-fold
 176 cross-validation, the parameters $n_{tree} = 2000$ and $m_{try} = 4$ are selected.

177 *Artificial neural network*

178 As the name suggested, this is a connectionist system which is inspired by
 179 the biological neural networks. It is also commonly known as the multilayer
 180 perceptron (MLP). A standard neural network consists of many connected
 181 nodes called neural, constructing the input, hidden and output layers. Each
 182 neuron produces a sequence of real-value activation. The input values are
 183 multiplied by the synaptic weights, which present the strength of the con-
 184 nection. The sum of these products are fed to each neuron within the hidden
 185 layer via a typically non-linear real-valued activation function such as tanh
 186 or logistic [40, 41]. In case of a single hidden layer, the values are then fed
 187 into the output layer neural via the activation function, and predict the out-

188 put value for each instance. Figure 4 illustrates the fully connected artificial
 189 neural network. During the training process, MLPs employ backpropagation
 190 techniques to minimize the sum of squared errors [42].

191 In this paper, we investigate the fully connected feed-forward neural net-
 192 work with one hidden layer; the inputs are parity, DIM, ..., NE; and the
 193 output is the milk yield. The R package 'neuralnet' [43] is used to implement
 194 the data in our study. To avoid overfitting the training data, heuristically,
 195 the optimum network consisted of 4 neurons in the hidden layer is used [9].
 196 The resilient backpropagation with weight backtracking is applied to train
 197 the data. The logistic function in (1) is carried out as the activation function:

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

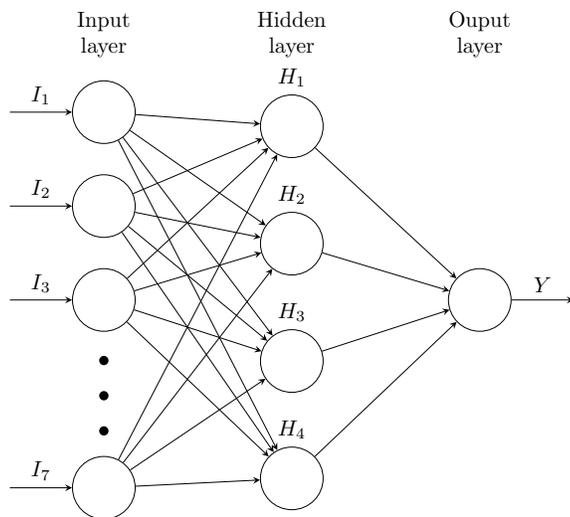


Figure 4: Artificial neural network with one hidden layer.

198 4. Prediction performance and discussion

199 In order to evaluate the prediction performance of the multiple linear
 200 regression (MLR) with elastic regression and the machine learning algorithms
 201 on this dataset; for each cow, the training set is the dataset excluding the data
 202 of one individual. The trained model is then used to predict the production
 203 of the excluded dairy cow. Moreover, the autoregressive version of these

204 methods are also investigated in this paper. The evaluation criteria chosen
205 in this study include: Root Mean Squared Error (RMSE), Mean Absolute
206 Error (MAE) and the variance of residual error, σ^2 . In addition, we also
207 compare the computational cost of each model to each other.

208 The computer used in this study was a MacBook Pro with Intel core i7 2.5
209 GHz and 16 G 1600 MHz DDR3. Table 5 and Figure 5 present the RMSE,
210 the MAE and the variance σ^2 of the elastic regression, SVR, random forest,
211 neural network forecasts against dataset of 36 individual cows in case of no
212 autoregression. The maximum and minimum RMSE values are 5.16 and
213 1.56 for the MLR, 4.61 and 1.44 for the SVR, 5.77 and 1.46 for the random
214 forest, 4.75 and 1.46 for the neural network. Table 6 shows the average
215 errors of each model for all 36 individual cows. In general, all the machine
216 algorithms mostly outperform the MLR. The random forest and SVR show
217 the most favorable results, and random forest model is more accurate in term
218 of RMSE and MAE. Moreover, in Table 7, the random forest can compute
219 the internal estimates of variable importance (in percentage). Similar to the
220 MLR, starch is the most importance variable according to the random forest
221 algorithm.

PLEASE PUT THE TABLE 5 HERE

PLEASE PUT THE FIGURE 5 HERE

PLEASE PUT THE TABLES 6, 7 HERE

222 As shown in Table 8, the MLR has the least training time (in seconds)
223 due to its simplicity, while the neural network model has the most expensive
224 computing. The SVR has a substantial better computational time than the
225 random forest, and is not much different from the MLR in computational
226 cost. Therefore, in term of both accuracy and computational cost, the SVR
227 gives the most sufficient result.

PLEASE PUT THE TABLE 8 HERE

228 A nonlinear autoregressive exogenous (NARX) model was applied to milk
229 production forecasting at herd level in the study by Murphy et al. [9]. In that
230 study, the training data consists of daily herd milk yield, days in milk and
231 number of cows milked, and the NARX was shown to be the most effective
232 milk-production model. In our study, the autoregressive version of the afore-
233 mentioned models is also considered. The autoregressive models applied in

234 our experiment have an order of one. In particular, the record in the previous
235 week is added into the prediction variables:

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

236 where y_t is the milk production record, and $\{u_1, u_2, \dots, u_p\}$ are the other
237 prediction variables. Table 9 and Figure 6 present the errors of the autore-
238 gressive version of all four forecasting models against dataset of 36 individual
239 cows. In all cases, the autoregressive approach significantly improves the ac-
240 curacy of all prediction models. For example, considering individual cow ID
241 7, the RMSEs of four models without autoregression are 2.44, 2.22, 2.81 and
242 2.67, respectively; with autoregression, the errors decreased to 1.88, 1.89,
243 2.35 and 1.80, respectively. However, considering the cow number 35, we get
244 more error with the autoregressive models, this can be caused by the status
245 of that individual (e.g. health problem). Therefore, milk yield forecasting
246 could be applied in monitoring health conditions [14]. In average, Table 10
247 show an improvement in accuracy compared to the model without autore-
248 gression. Moreover, as shown in Table 11, the internal estimates of variable
249 importance computed by random forest show that the information in the
250 past is essentially important (62.78%), starch is still an important variable
251 (14.81%) compared to the rest.

PLEASE PUT THE TABLE 9 HERE

PLEASE PUT THE FIGURE 6 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 may be suitable to commercial applications.

PLEASE PUT THE TABLE 12 HERE

252

253 **5. Concluding remarks**

254 This is a study on the small scale (36 milking cows) in Brittany, France.
255 The correlation between fat and protein content and milk yield with the
256 collected data has indicated the decrease of the fat and protein content as
257 milk yield increases to 20 (kg/day). On this dataset, the importance of the
258 chemical composition of nutrition on the milk production of dairy cattle has
259 shown the significant effect of the quantity of starch supply through the diet
260 on milk production level.

261 In this study, we compare the performance of the linear regression mod-
262 els and machine learning models on forecasting milk production at individ-
263 ual level. The hyper parameters are tuned with cross-validation criteria.
264 The autoregressive models are shown to be significant better than the non-
265 autoregressive approaches. When the past is considered, the information
266 from the previous observation considerably improves the prediction accu-
267 racy. The support vector machine is the most efficient method among the
268 other models in term of both prediction accuracy and computational cost.
269 The result indicates the possibility of practical application on the small scale
270 farm with a small number of dairy cows. However, the autoregressive models
271 require the previous observation, then the non-autoregressive approaches are
272 more practical when past observations are not available, or a far prediction is
273 considered. Further research on other kind of dairy cows with more intensive
274 observation is required to investigate the potential of using these models in
275 health monitoring on an individual cow level with high accuracy.

276 **Acknowledgments**

277 This research activity have been financed by Conseil regional Bretagne
278 and FEDER Bretagne within the project NUTGEN of the Université de
279 Bretagne Sud.

280 **References**

- 281 [1] P. Wood, Algebraic model of the lactation curve in cattle, *Nature* 216
282 (1967) 164–165.
- 283 [2] T. Ali, L. Schaeffer, Accounting for covariances among test day milk
284 yields in dairy cows, *Canadian Journal of Animal Science* 67 (1987)
285 637–644.

- 286 [3] J. Wilmink, Adjustment of lactation yield for age at calving in relation
287 to level of production, *Livestock Production Science* 16 (1987) 321–334.
- 288 [4] L. Schaeffer, Application of random regression models in animal breed-
289 ing, *Livestock Production Science* 86 (2004) 35–45.
- 290 [5] A. Silvestre, A. Martins, V. Santos, M. Ginja, J. Colaço, Lactation
291 curves for milk, fat and protein in dairy cows: A full approach, *Livestock*
292 *Science* 122 (2009) 308–313.
- 293 [6] S. Adediran, D. Ratkowsky, D. Donaghy, A. Malau-Aduli, Comparative
294 evaluation of a new lactation curve model for pasture-based holstein-
295 friesian dairy cows, *Journal of dairy science* 95 (2012) 5344–5356.
- 296 [7] M. Mellado, J. Flores, A. De Santiago, F. Veliz, U. Macías-Cruz,
297 L. Avendaño-Reyes, J. García, Extended lactation in high-yielding hol-
298 stein cows: Characterization of milk yield and risk factors for lactations
299 > 450 days, *Livestock Science* 189 (2016) 50–55.
- 300 [8] J. O. Lehmann, L. Mogensen, T. Kristensen, Extended lactations in
301 dairy production: Economic, productivity and climatic impact at herd,
302 farm and sector level, *Livestock science* 220 (2019) 100–110.
- 303 [9] M. Murphy, M. O’Mahony, L. Shalloo, P. French, J. Upton, Comparison
304 of modelling techniques for milk-production forecasting, *Journal of dairy*
305 *science* 97 (2014) 3352–3363.
- 306 [10] F. Zhang, M. D. Murphy, L. Shalloo, E. Ruelle, J. Upton, An automatic
307 model configuration and optimization system for milk production fore-
308 casting, *Computers and Electronics in Agriculture* 128 (2016) 100–111.
- 309 [11] S. Nickerson, Milk production: Factors affecting milk composition, in:
310 *Milk quality*, Springer, 1995, pp. 3–24.
- 311 [12] I. Harder, E. Stamer, W. Junge, G. Thaller, Lactation curves and model
312 evaluation for feed intake and energy balance in dairy cows, *Journal of*
313 *Dairy Science* (2019).
- 314 [13] J. Boerman, S. Potts, M. VandeHaar, M. Allen, A. Lock, Milk pro-
315 duction responses to a change in dietary starch concentration vary by
316 production level in dairy cattle, *Journal of dairy science* 98 (2015) 4698–
317 4706.

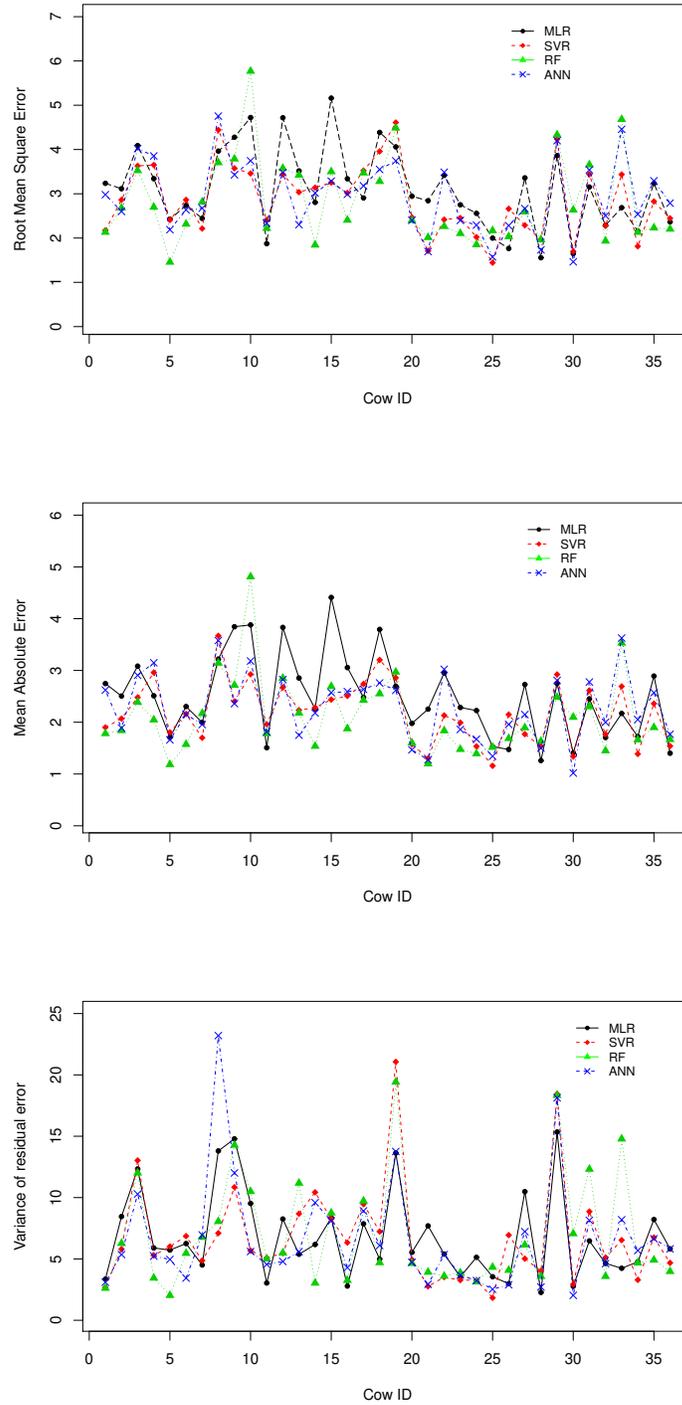
- 318 [14] F. Andersen, O. Østerås, O. Reksen, Y. T. Gröhn, Mastitis and the
319 shape of the lactation curve in norwegian dairy cows, *Journal of dairy*
320 *research* 78 (2011) 23–31.
- 321 [15] D. B. Jensen, M. van der Voort, H. Hogeveen, Dynamic forecasting of
322 individual cow milk yield in automatic milking systems, *Journal of dairy*
323 *science* 101 (2018) 10428–10439.
- 324 [16] F. Zhang, J. Upton, L. Shalloo, M. Murphy, Effect of parity weight-
325 ing on milk production forecast models, *Computers and Electronics in*
326 *Agriculture* 157 (2019) 589–603.
- 327 [17] J. Van Bebber, N. Reinsch, W. Junge, E. Kalm, Monitoring daily milk
328 yields with a recursive test day repeatability model (kalman filter), *Jour-*
329 *nal of dairy science* 82 (1999) 2421–2429.
- 330 [18] M. Vermorel, Energy: the feed unit systems, in: *Ruminant nutrition:*
331 *recommended allowances and feed tables*, INRA Publications, Paris,
332 1989, p. 28.
- 333 [19] S. A. Kadi, F. Djellal, M. Berchiche, Caractérisation de la conduite ali-
334 mentaire des vaches laitières dans la région de tizi-ouzou, algérie, *Live-*
335 *stock Research for rural development* 19 (2007).
- 336 [20] G. Pulina, A. Nudda, G. Battaccone, A. Cannas, Effects of nutrition on
337 the contents of fat, protein, somatic cells, aromatic compounds, and un-
338 desirable substances in sheep milk, *Animal Feed Science and Technology*
339 131 (2006) 255–291.
- 340 [21] R. Emery, Milk fat depression and the influence of diet on milk compo-
341 sition., *The Veterinary Clinics of North America. Food Animal Practice*
342 4 (1988) 289–305.
- 343 [22] U. Meyer, M. Everinghoff, D. Gädeken, G. Flachowsky, Investigations
344 on the water intake of lactating dairy cows, *Livestock production science*
345 90 (2004) 117–121.
- 346 [23] G. Pulina, N. Macciotta, A. Nudda, Milk composition and feeding in
347 the italian dairy sheep, *Italian Journal of Animal Science* 4 (2005) 5–14.

- 348 [24] V. Olori, S. Brotherstone, W. Hill, B. McGuirk, Fit of standard models
349 of the lactation curve to weekly records of milk production of cows in a
350 single herd, *Livestock Production Science* 58 (1999) 55–63.
- 351 [25] A. E. Hoerl, R. W. Kennard, Ridge regression: Biased estimation for
352 nonorthogonal problems, *Technometrics* 12 (1970) 55–67.
- 353 [26] J. O. Ogutu, T. Schulz-Streeck, H.-P. Piepho, Genomic selection using
354 regularized linear regression models: ridge regression, lasso, elastic net
355 and their extensions, *BMC Proceedings* 6 (2012) S10.
- 356 [27] R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal*
357 *of the Royal Statistical Society: Series B (Methodological)* 58 (1996)
358 267–288.
- 359 [28] H. Zou, T. Hastie, Regularization and variable selection via the elastic
360 net, *Journal of the royal statistical society: series B (statistical method-*
361 *ology)* 67 (2005) 301–320.
- 362 [29] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for general-
363 ized linear models via coordinate descent, *Journal of statistical software*
364 33 (2010) 1.
- 365 [30] A. Cabrita, R. Bessa, S. Alves, R. Dewhurst, A. Fonseca, Effects of
366 dietary protein and starch on intake, milk production, and milk fatty
367 acid profiles of dairy cows fed corn silage-based diets, *Journal of Dairy*
368 *Science* 90 (2007) 1429–1439.
- 369 [31] C. Kamphuis, H. Mollenhorst, A. Feelders, D. Pietersma, H. Hogeveen,
370 Decision-tree induction to detect clinical mastitis with automatic milk-
371 ing, *Computers and Electronics in Agriculture* 70 (2010) 60–68.
- 372 [32] K. Saruta, Y. Hirai, K. Tanaka, E. Inoue, T. Okayasu, M. Mitsuoka,
373 Predictive models for yield and protein content of brown rice using sup-
374 port vector machine, *Computers and electronics in agriculture* 99 (2013)
375 93–100.
- 376 [33] B. Barrett, I. Nitze, S. Green, F. Cawkwell, Assessment of multi-
377 temporal, multi-sensor radar and ancillary spatial data for grasslands
378 monitoring in ireland using machine learning approaches, *Remote Sens-*
379 *ing of Environment* 152 (2014) 109–124.

- 380 [34] P. Shine, M. D. Murphy, J. Upton, T. Scully, Machine-learning algo-
381 rithms for predicting on-farm direct water and electricity consumption
382 on pasture based dairy farms, *Computers and electronics in agriculture*
383 150 (2018) 74–87.
- 384 [35] A. J. Smola, B. Schölkopf, A tutorial on support vector regression,
385 *Statistics and computing* 14 (2004) 199–222.
- 386 [36] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, C.-C.
387 Chang, C.-C. Lin, M. D. Meyer, Package ‘e1071’, *The R Journal* (2019).
- 388 [37] L. Breiman, Random forests, *Machine learning* 45 (2001) 5–32.
- 389 [38] A. Liaw, M. Wiener, et al., Classification and regression by randomfor-
390 est, *R news* 2 (2002) 18–22.
- 391 [39] L. Breiman, *Classification and regression trees*, Routledge, 2017.
- 392 [40] P. Bickel, P. Diggle, S. Fienberg, U. Gather, I. Olkin, S. Zeger, *Springer*
393 *Series in Statistics*, Springer, 2009.
- 394 [41] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural*
395 *networks* 61 (2015) 85–117.
- 396 [42] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hub-
397 bard, L. D. Jackel, Backpropagation applied to handwritten zip code
398 recognition, *Neural computation* 1 (1989) 541–551.
- 399 [43] F. Günther, S. Fritsch, neuralnet: Training of neural networks, *The R*
400 *journal* 2 (2010) 30–38.

Cow ID	MLR			SVR			Random forest			Neural network		
	RMSE	MAE	σ^2	RMSE	MAE	σ^2	RMSE	MAE	σ^2	RMSE	MAE	σ^2
1	3.24	2.75	3.35	2.17	1.90	2.68	2.13	1.78	2.62	2.98	2.62	3.10
2	3.11	2.51	8.45	2.86	2.07	5.80	2.68	1.84	6.27	2.60	1.89	5.36
3	4.09	3.08	12.33	3.63	2.48	13.03	3.52	2.39	12.01	4.02	2.91	10.29
4	3.34	2.51	5.89	3.65	2.96	5.29	2.70	2.05	3.44	3.85	3.15	5.23
5	2.42	1.71	5.72	2.41	1.81	6.01	1.46	1.18	2.03	2.19	1.66	4.96
6	2.74	2.30	6.25	2.86	2.16	6.86	2.32	1.57	5.46	2.64	2.15	3.44
7	2.44	1.99	4.51	2.22	1.70	4.87	2.81	2.17	6.76	2.67	1.96	6.96
8	3.96	3.22	13.80	4.44	3.67	7.10	3.70	3.14	8.05	4.75	3.58	23.20
9	4.28	3.84	14.79	3.58	2.40	10.85	3.79	2.72	14.28	3.43	2.36	12.01
10	4.72	3.88	9.52	3.46	2.93	5.67	5.77	4.81	10.5	3.75	3.18	5.60
11	1.87	1.51	3.04	2.41	1.96	5.08	2.22	1.78	5.01	2.33	1.82	4.56
12	4.72	3.83	8.25	3.44	2.67	5.47	3.58	2.85	5.48	3.48	2.82	4.79
13	3.52	2.85	5.38	3.04	2.23	8.68	3.42	2.18	11.18	2.3	1.75	5.59
14	2.81	2.26	6.17	3.14	2.28	10.43	1.84	1.54	3.04	3.02	2.18	9.59
15	5.16	4.41	8.36	3.25	2.44	8.26	3.50	2.69	8.75	3.28	2.57	8.09
16	3.34	3.06	2.79	3.02	2.51	6.33	2.41	1.88	3.26	2.99	2.59	4.29
17	2.91	2.48	7.86	3.52	2.74	9.47	3.47	2.43	9.72	3.18	2.63	8.91
18	4.38	3.79	4.99	3.96	3.20	7.23	3.28	2.55	4.70	3.55	2.76	6.10
19	4.06	2.70	13.63	4.61	2.86	21.07	4.49	2.97	19.42	3.74	2.62	13.76
20	2.94	1.98	5.54	2.47	1.58	4.93	2.40	1.6	4.64	2.40	1.47	4.73
21	2.84	2.25	7.70	1.71	1.30	2.77	2.01	1.2	3.92	1.70	1.27	2.93
22	3.42	2.95	5.39	2.42	2.13	3.52	2.26	1.84	3.6	3.49	3.02	5.35
23	2.75	2.28	3.55	2.45	1.99	3.27	2.10	1.48	3.87	2.40	1.86	3.67
24	2.56	2.23	5.13	2.02	1.53	3.27	1.85	1.39	3.13	2.29	1.67	3.23
25	2.00	1.53	3.55	1.44	1.16	1.83	2.17	1.52	4.33	1.57	1.34	2.53
26	1.76	1.47	2.98	2.66	2.15	6.94	2.03	1.69	4.08	2.28	1.96	2.89
27	3.36	2.73	10.48	2.29	1.77	5.02	2.59	1.9	6.15	2.67	2.15	7.22
28	1.56	1.26	2.27	1.97	1.54	4.02	1.96	1.63	3.58	1.73	1.50	2.71
29	3.86	2.75	15.35	4.23	2.92	18.45	4.33	2.48	18.33	4.20	2.81	18.13
30	1.65	1.41	2.77	1.70	1.34	2.93	2.64	2.10	7.06	1.46	1.02	2.03
31	3.15	2.45	6.47	3.44	2.61	8.86	3.66	2.30	12.32	3.54	2.78	8.14
32	2.29	1.71	4.63	2.28	1.77	5.12	1.93	1.45	3.58	2.50	2.00	4.62
33	2.69	2.17	4.24	3.44	2.69	6.53	4.68	3.53	14.78	4.46	3.63	8.19
34	2.16	1.72	4.76	1.81	1.39	3.29	2.13	1.66	4.68	2.54	2.05	5.69
35	3.24	2.89	8.21	2.83	2.36	6.74	2.23	1.90	4.91	3.29	2.57	6.65
36	2.36	1.40	5.82	2.45	1.54	4.68	2.20	1.67	3.98	2.79	1.77	5.82

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



21
 Figure 5: Root Mean Squared Error (above), Mean Absolute Error (middle) and variance of residual error (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
σ^2	6.776	6.732	6.915	6.677

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

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

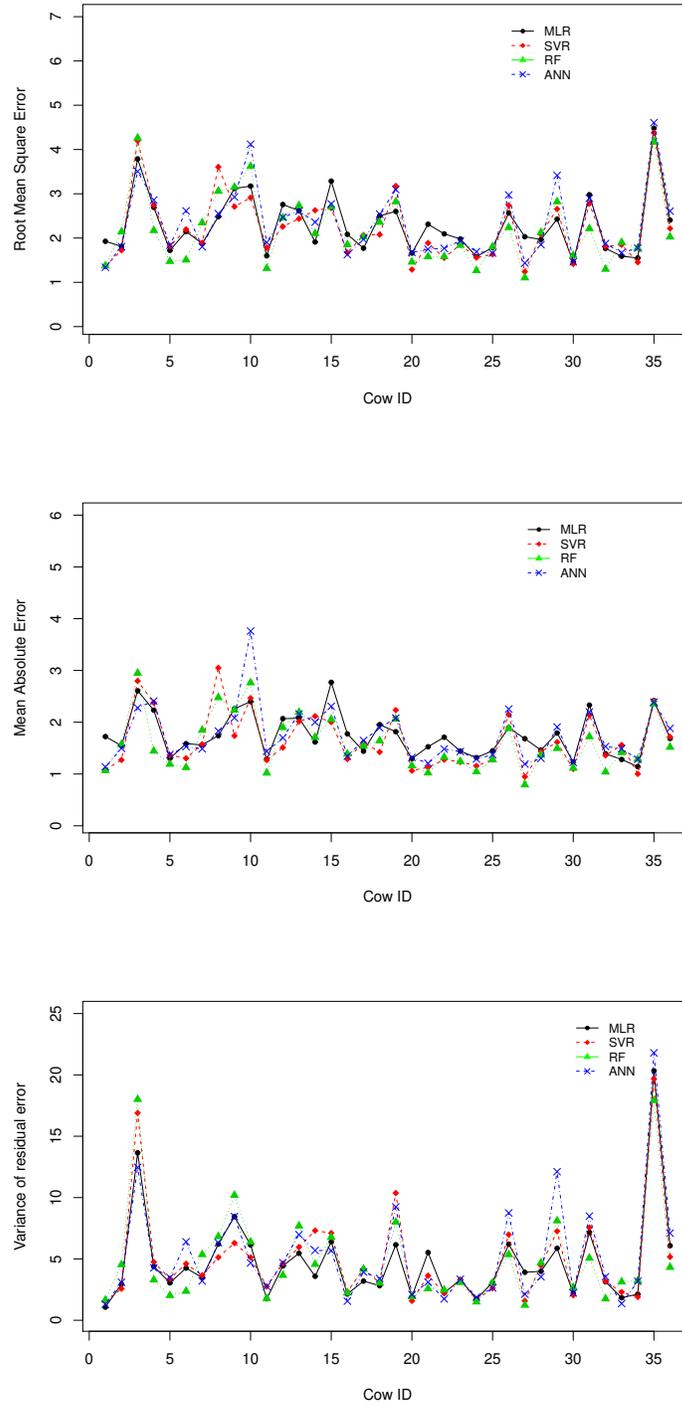
Table 7: Average of variable importance estimated by random forest (in %).

	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.

Cow ID	MLR			SVR			Random forest			Neural network		
	RMSE	MAE	σ^2	RMSE	MAE	σ^2	RMSE	MAE	σ^2	RMSE	MAE	σ^2
1	1.93	1.72	1.07	1.36	1.09	1.55	1.37	1.07	1.65	1.34	1.14	1.30
2	1.81	1.55	2.94	1.72	1.27	2.57	2.14	1.57	4.54	1.81	1.49	3.10
3	3.79	2.61	13.65	4.20	2.80	16.9	4.26	2.95	18.01	3.51	2.28	12.47
4	2.69	2.23	4.39	2.75	2.38	4.75	2.17	1.45	3.30	2.86	2.41	4.27
5	1.72	1.30	3.05	1.81	1.37	3.39	1.47	1.19	2.02	1.84	1.38	3.49
6	2.15	1.59	4.25	2.20	1.30	4.60	1.51	1.13	2.37	2.61	1.53	6.40
7	1.88	1.56	3.48	1.89	1.57	3.66	2.35	1.85	5.36	1.80	1.49	3.20
8	2.48	1.74	6.21	3.61	3.05	5.13	3.06	2.48	6.83	2.55	1.83	6.32
9	3.12	2.26	8.45	2.71	1.74	6.30	3.15	2.24	10.19	2.93	2.08	8.45
10	3.17	2.40	6.13	2.91	2.47	5.13	3.62	2.76	6.39	4.12	3.76	4.67
11	1.60	1.28	1.75	1.78	1.27	2.76	1.32	1.02	1.77	1.90	1.44	2.77
12	2.76	2.07	4.43	2.26	1.51	4.62	2.46	1.90	3.67	2.47	1.70	4.69
13	2.62	2.08	5.46	2.44	2.01	5.96	2.74	2.19	7.69	2.61	2.16	6.98
14	1.91	1.62	3.58	2.63	2.12	7.31	2.10	1.71	4.57	2.36	2.00	5.69
15	3.29	2.77	6.39	2.68	2.00	7.11	2.71	2.06	6.78	2.77	2.31	5.67
16	2.08	1.78	2.14	1.67	1.29	2.32	1.85	1.39	2.23	1.63	1.35	1.56
17	1.77	1.44	3.19	2.07	1.61	4.20	2.03	1.54	4.17	1.97	1.65	3.93
18	2.50	1.95	2.82	2.08	1.43	3.06	2.36	1.64	3.04	2.54	1.89	3.40
19	2.60	1.82	6.15	3.18	2.23	10.37	2.82	2.06	8.01	3.10	2.08	9.24
20	1.66	1.30	1.93	1.29	1.06	1.58	1.46	1.16	1.93	1.67	1.32	2.10
21	2.31	1.53	5.52	1.89	1.13	3.63	1.58	1.02	2.58	1.75	1.21	3.11
22	2.10	1.71	2.23	1.55	1.28	2.22	1.58	1.33	2.47	1.77	1.48	1.74
23	1.98	1.44	3.31	1.85	1.23	3.28	1.84	1.24	3.08	1.94	1.43	3.34
24	1.58	1.32	1.73	1.56	1.16	1.65	1.27	1.05	1.50	1.69	1.29	1.88
25	1.79	1.45	3.09	1.63	1.28	2.62	1.81	1.28	3.05	1.66	1.37	2.63
26	2.57	1.89	6.20	2.74	2.15	6.98	2.24	1.88	5.35	2.97	2.25	8.75
27	2.03	1.68	3.91	1.24	0.95	1.59	1.10	0.79	1.23	1.43	1.19	2.12
28	1.97	1.46	3.99	2.08	1.43	4.47	2.13	1.37	4.69	1.85	1.30	3.54
29	2.43	1.79	5.87	2.66	1.62	7.26	2.82	1.5	8.11	3.42	1.91	12.11
30	1.46	1.23	2.19	1.42	1.10	2.04	1.61	1.11	2.68	1.46	1.22	2.23
31	2.98	2.33	7.16	2.77	2.14	7.56	2.21	1.72	5.07	2.90	2.20	8.49
32	1.77	1.39	3.14	1.81	1.36	3.17	1.30	1.04	1.75	1.88	1.53	3.55
33	1.59	1.28	1.84	1.85	1.56	2.30	1.90	1.42	3.13	1.67	1.49	1.36
34	1.55	1.14	2.13	1.45	1.00	1.90	1.77	1.29	3.23	1.77	1.28	3.18
35	4.48	2.42	20.34	4.38	2.41	19.68	4.17	2.35	17.92	4.61	2.39	21.80
36	2.40	1.69	6.08	2.22	1.72	5.17	2.03	1.52	4.32	2.61	1.88	7.10

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



24
 Figure 6: Root Mean Squared Error (above), Mean Absolute Error (middle) and variance of residual error (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
σ^2	4.727	4.966	4.852	5.185

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.