



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.



Forecasting Regional Milk Production Quantity: A Comparison of Regression Models and Machine Learning

by Dominik Baaken and Sebastian Hess

*Copyright 2021 by Dominik Baaken and Sebastian Hess. All rights reserved.
Readers may make verbatim copies of this document for non-commercial purposes
by any means, provided that this copyright notice appears on all such copies.*

Forecasting Regional Milk Production Quantity: A Comparison of Regression Models and Machine Learning

Dominik Baaken and Sebastian Hess¹

31st International Conference of Agricultural Economists

August 17-31, 2021 – online

Abstract

Since the abolition of the European milk quota in 2015, there has been increasing demand from market participants for instruments to forecast regional milk quantities, but as yet there are none. Current approaches for milk production forecasting are mostly based on predicting the lactation curve of individual dairy cows. This paper compares six modelling approaches from different areas of machine learning (ML) and linear regression (OLS) that combine different blocks of variables to include temporal trends, direct and indirect weather effects, and price events in the prediction. For a prediction period of 21 months, a maximum coefficient of determination of 0.92 by ML methods and a maximum of 0.77 by OLS regressions can be obtained for north German farms. The comparison of ML algorithms reveals a difference in the models, particularly in relation to training speeds.

Keywords: Dairy market; milk production; predictive models; machine learning.

¹ Agricultural Markets Department, University of Hohenheim, Germany;
contact: dominik.baaken@uni-hohenheim.de

1. Introduction

Germany is the fourth largest milk-producing country in the world in terms of volume (FAOSTAT, 2020). In the last 10 years, the volume of milk produced in the country has increased by about 11.7 % (BLE, 2020a). Milk production plays a significant role in the economic sector of agriculture, with the dairy sector accounting for about 18.9 % of the value of goods produced (BMEL, 2020a). During this period the milk price has been very volatile, with a fluctuation range of 20.6 to 42.4 ct/l (BMEL, 2020b). Since the end of the milk quota in Europe in 2015, it has become much easier for farmers to expand their production capacities. One consequence of this is a lack of information among market participants about the volume of milk that will be produced in the months ahead.

Some dairies try to estimate the milk volume on the basis of voluntary self-reporting and self-assessment by farmers. However, it is not sufficient to simply ask farmers about their production intentions in order to build up a forecast index for predicting milk supply volumes because not all production-relevant influences can be assessed equally correctly by all farms at any given time (Larcher *et al.*, 2015). Moreover, from an economic point of view, strategic responses are to be expected in the case of self-reporting. The challenge in forecasting milk quantities lies in the dependence on slow, biological processes (Hoehl and Hess, 2021). These complicate farmers' ability to fine-tune and adjust their production volumes directly. For many years, econometric models have been estimated based on the lactation curve of individual dairy cows and thus for an entire farm (Adediran *et al.*, 2012). However, predictive models for a larger number of farms or even whole regions do not yet exist or are not publicly available. In this context, machine learning (ML) methods could offer a way of ensuring practice-relevant forecasting based on very large datasets. Instead of building and applying a model based on existing theory, ML reads patterns from existing data structures from which inferences about future developments can be drawn (Mullainathan and Spiess, 2017). The focus of ML methods

is therefore less on the correlations within a historical dataset and more on prediction outside the existing data (Storm *et al.*, 2019).

This paper applies different forecasting approaches of ML to predict milk quantities and compares them, using a conventional regression model as a reference. It identifies which type of ML algorithms can be applied to predict regional milk production and how they differ from each other. In addition, the combination and configuration of trend, weather and price variables relevant for the forecast are of interest. The data structure and preparation are described below, followed by an introduction to the models and methods applied. Section 4 presents and discusses the results of the algorithms before conclusions are drawn about future forecasting models.

2. Theoretical background

While economists usually describe aggregate dairy farm output in terms of a production technology with milk and meat output as a function of capital, labor and land inputs, the discipline of animal sciences has traditionally followed a different path: Production on dairy farms is particularly determined by the lactation curve of individual dairy cows. This can be illustrated as a mathematical formula to represent the course of milk production over time. Since Wood (1967) first established the *gamma* function, a variety of models have been developed to represent biological processes as accurately as possible. The course of the lactation curve is mainly based on the genetic material of the animals under consideration (Brügemann *et al.*, 2011). Thus, a similar curve per animal can be observed over several lactation phases. Deviations from the original curve can be caused by various factors such as grazing conditions (Baudracco *et al.*, 2011), seasonality (Wood, 1969), feeding conditions (Kolver and Muller, 1998), diseases (Collard *et al.*, 2000) and other disturbances (Teklerli *et al.*, 2000) such as heat. While farmers and their herd management can control most factors, feed conditions depend on weather as an additional component: Several studies show the influence of weather on pasture

and forage quality (Parsons *et al.*, 2019; Schlenker and Roberts, 2009). Therefore, weather can be expected to influence milk production in the short and medium term. In addition, livestock conditions are also related to weather. A combination of temperature, relative humidity, solar radiation, wind speed and precipitation can cause heat stress in animals, leading to reduced milk production in terms of quantity and quality (Mader *et al.*, 2006). High temperatures can also cause increased mortality in dairy cows (Vitali *et al.*, 2015).

Farmers' herd management is modelled in different ways in the literature. Dual production functions model the quantity of milk farms produce as a relationship between output quantities and input and output prices. This is based on indirect cost or profit functions that are designed to maximise profits. For example, Blayney and Mittelhammer (1990) use a profit function of a generalised Box-Cox form to calculate price elasticities of the quantity of produced milk. Thus, it is possible to simulate the amount by which the milk supply changes in relation to price changes. Bouamra-Mechemache *et al.* (2008) also use a dual approach and determine short and medium-term price elasticities for the quantity of milk produced using a normalised quadratic profit function. Both models illustrate the relationship between milk production and input and output prices. Farmers act as a profit maximiser in this relationship, making management decisions to adjust their milk production at given prices.

The amount of milk produced in a region is the sum of the amount of milk produced by the farms located there. In the last 20 years, the number of farms in Germany has fallen by more than half, but the quantity of milk produced has increased by 14.4 % in the same period (BLE, 2020a). This is caused by ongoing structural change, with smaller and less profitable farms ceasing production due to various factors such as economics of size and scale (Hallam, 1991) and natural conditions and agricultural prices in specific regions (Neuenfeldt *et al.*, 2019). In contrast, larger and highly professionalised dairy farms can significantly increase their production through increased remounting and expansion of their facilities.

Figure 1 shows the relations of the determinants of milk production outlined above. For the forecasting models presented in this paper, the influencing factors are grouped into three main components: weather, time and price variables. It is assumed that milk production in a region k is a function depending on the production function on farm i with cows j kept and their lactation curve. This results in a function depending on herd performance, herd management, market prices and other factors. This simplified representation allows complex on-farm interactions and macroeconomic developments to be illustrated by so called time variables as a trend in milk production.

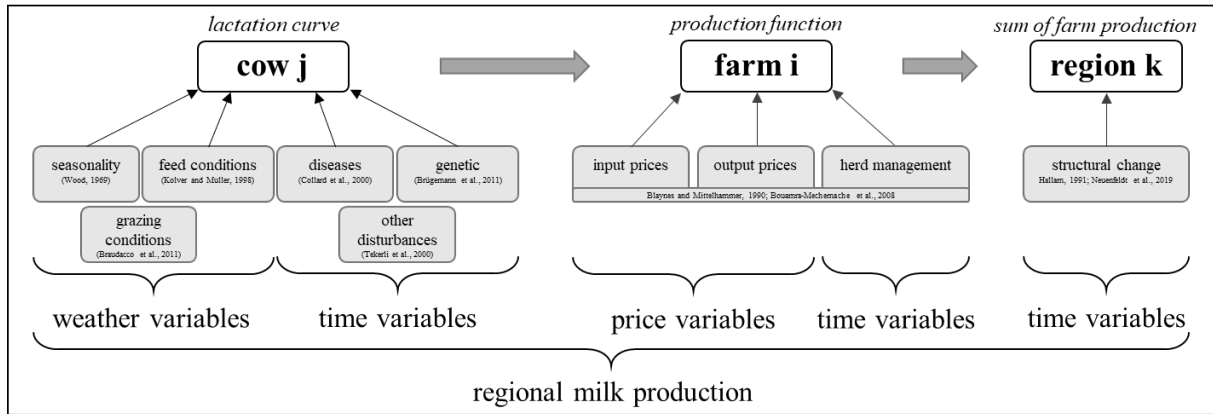


Figure 1. Schematic structure of milk production

An exact definition of the variable groups used with the specific explanation of individual variables is given below.

3. Data collection

The dataset used refers to monthly observations in the period 09/2009 to 09/2019, and thus falls roughly within the timeframe when the abolition of the milk quota had already been decided and prepared for by an annual increase in quota volumes (“soft landing”). Variables of individual farm milk production as well as weather and price data are included in the models.

The data on milk production are provided by the “Landeskontrollverband Niedersachsen e.V.” (LKV). The association compiles a monthly milk performance test for all member farms in Lower Saxony (a region in northern Germany) and provides this data for 49 randomly selected

dairy farms. This generates data points for about 7,500 dairy cows in Lower Saxony that are aggregated by an average value for the respective farms. Thus, a single monthly value is obtained by calculating the mean milk production of the cows milked on each farm. This value is multiplied by 30.5 days and the number of milked animals per farm and month, corresponding to the average amount of milk produced per farm and month. The “energy-corrected milk yield” (ECM) is considered, calculated according to Sjaunja *et al.* (1990). The number of animals kept per farm varies from fewer than 30 to more than 500 animals. The mean is about 150 cows per farm, with the majority of farms (about 60 %) being in the size class of farms with 50-199 dairy cows. This roughly corresponds to the distribution of size classes in Lower Saxony. However, farms with more than 200 cows are overrepresented in the data with a share of almost 30 % compared with just under 10 % of farms in Lower Saxony (LSN, 2020). However, this is helpful in terms of forecasting as structural change is ongoing and the share of the size classes is continuously shifting towards larger farms (BLE, 2020b).

As shown above, it must be assumed that milk production in a region is largely determined by short and long-term climatic influences (Hoehl and Hess, 2021). This particularly affects the quality of basic feed and the physiology of the animals. It would therefore appear necessary to include regional weather and climate data in the forecast. For weather variables, the “COSMO Regional Reanalysis” database of the German Weather Service (DWD) is used. In these numerical weather models, observations and analyses are combined so that a large number of parameters are available for a grid of about 6 km x 6 km resolution.

Milk prices are provided by the Federal Ministry of Agriculture and Food (BMEL) and refer to conventionally produced cow’s milk in Lower Saxony at actual fat and protein content. For feed prices, the index for purchase prices of agricultural enterprises determined by the Federal Statistical Office is used.

For further analysis, the dataset is divided into a training set and a test set. This allows the models and algorithms to learn the patterns of the data in the training set. In a second step, the

prediction precision of the algorithms can be evaluated by applying them to the test set. A typical split of about 80 % training data and 20 % test data is used, such that all data prior to 2018 is considered as the training set and from 01/2018 is considered as the test period.

4. Methods

This section presents the different models used to predict milk production in the region of Lower Saxony. The basic model is a multiple linear regression based on the ordinary least squares (OLS) estimator. Compared with this basic model, a number of machine learning (ML) models are set up to compare the difference to the econometric approach and between the different ML algorithms for the prediction of milk production.

The considered farms can be divided into two classes based on their annual growth rate of milk produced. To account for the different types of farms, an indicator function is introduced before the algorithms are applied to the dataset. This indexed distinction allows the algorithms to learn the patterns in the data more effectively and generalise to new data. The indicator function is defined as follows:

$$I_{GR}(x) = \begin{cases} 1 & \text{if } x_i \leq GR \\ 0 & \text{if } x_i > GR \end{cases} \quad (1)$$

where x_i is the average growth rate of milk production for farm i and $GR = Q_{0.8}(x)$. By making this split based on the 0.8 quantile, for all model specifications k a sub-model is formed for group a with $x_i > GR$ and for group b with $x_i \leq GR$. In the results section, both groups are combined for each model k , resulting in a complete model.

The majority of farms (80 %) show an annual change in milk quantity of between approximately -1.5 % and 9 %. In contrast, the remaining farms are growing at an annual rates of around 14 %, with noticeable jumps in the monthly volume produced (Fig. 2). This suggests that these farms are growing not only through on-farm remounting, but through larger investments based on purchased animals. The farms can therefore be divided into two groups based on their growth

rate. The resulting different development in the produced milk volume over time is shown in Figure 2a. The trend line illustrates that the production volume increases more quickly in group a (rapidly growing farms) than in group b (moderately growing farms).

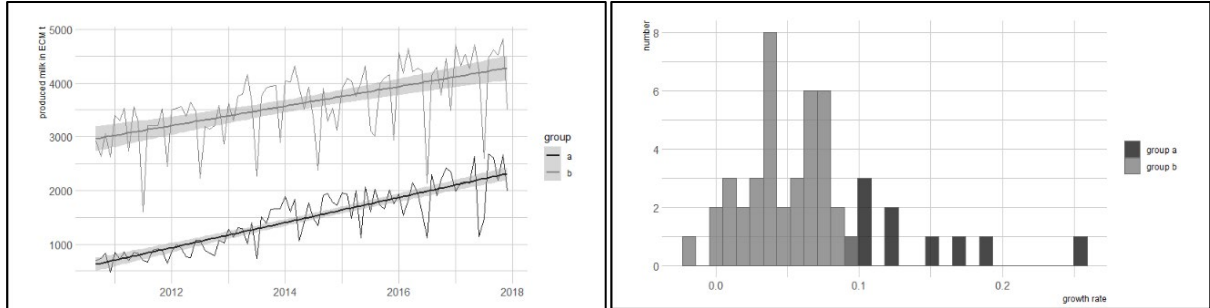


Figure 2a: Milk production over time

Figure. 2b: Histogram growth rates

4.1. Basic model: Forecast of monthly milk production amounts

The basic model assumes that the milk production of farm i can be explained by an interaction of economic and climatic conditions in month t and previous months (lags). Economic effects can be divided into price effects for inputs and outputs and farm-specific management factors. The influence of weather on milk yield can be divided into direct and indirect effects. Direct weather effects typically include heat stress for dairy cows and impairment of their metabolism (Hill and Wall, 2017). Indirect weather effects result from quality and quantity differences in feeding (Harder *et al.*, 2019).

The basic model can be described as follows:

$$y_{it} = T_t + W_{it} + P_t + \mu_i + \varepsilon_{it} \quad (2)$$

where y_{it} = average monthly quantity of milk [kg ECM] produced by farm i at time (month) t , T_t = time variables, W_{it} = weather variables, P_t = price variables, μ_i = dummy for farm i and ε_{it} = error term. The time variables are divided into date, year and factor variables for each month. The weather variables section contains ten different weather aspects at time t , which can be assigned to the coordinates of farm i . These are precipitation [mm], minimum temperature [°C], maximum temperature [°C], mean temperature [°C], soil temperature [°C],

relative humidity [%], sunshine duration [h], cloud cover [%], wind speed [km/h] and temperature-humidity index (calculated according to Mader *et al.* (2006)). The monthly milk prices for Lower Saxony [ct/l] and a feed price index at time t are used as price variables.

Table 1 shows the combination of variable groups for each model, presenting the differences between the models as an overview.

Table 1
Overview model specifications

Equation	Model	Weather	Interaction term weather and monthly dummy (MD)	Weather lags	Interaction term weather lags and monthly dummy (MD)	Prices	Dummy per farm
		W_{it}	$W_{it} * MD_m$	$W_{i,t-n}$	$W_{i,t-n} * MD_{m-n}$	P_t	μ_i
2	A	X				X	X
3	B		X			X	X
4	C				X	X	X
5	D		X		X	X	X
6	E	X	X	X	X	X	X

The basic model represents direct weather effects only. For the other models, an interaction term between the direct weather variables and a dummy MD_m is introduced for each month m with $m = \{January, \dots, December\}$. This interaction term is included in equation (3) in the model specification B:

$$y_{it} = T_t + W_{it} * MD_m + P_t + \mu_i + \varepsilon_{it} \quad (3)$$

To model indirect weather effects, lags in the weather variables are added to equation (3), while the direct effects term is omitted. This leads to the following equation:

$$y_{it} = T_t + W_{i,t-n} * MD_{m-n} + P_t + \mu_i + \varepsilon_{it} \quad (4)$$

where W_{t-n} = weather variables at time $t - n$ and MD_{m-n} = dummy for month $m - n$, where $n = \{1, \dots, 6\}$. All other variables are defined as in the basic model.

These two models of direct and indirect effects are also combined into one model as a combination of both effects, so that the model is constructed as follows:

$$y_{it} = T_t + W_{it} * MD_m + W_{i;t-n} * MD_{m-n} + P_t + \mu_i + \varepsilon_{it} \quad (5)$$

The final model combines all of the different model types into one model for all of the variables created, which is shown as follows:

$$y_{it} = T_t + W_{it} + W_{it} * MD_m + W_{i;t-n} + W_{i;t-n} * MD_{m-n} + P_t + \mu_i + \varepsilon_{it} \quad (6)$$

Lags in the price variables do not need to be included since price trends and price expectations are included in the farm-specific growth rate and the trend variable T .

4.2. Machine Learning Models

The model specifications presented in the previous section are applied to the dataset by means of different estimation methods, comparing conventional regressions² (OLS) with so-called “supervised learning” machine learning (ML) methods. In ML, the variables are called features. As there is a number of supervised ML approaches, the analysis here is limited to three different model types: regularised regression, tree-based methods and a neural network approach.

4.2.1. Regularised regression

Regularisation is based on the addition of a penalty term to a linear regression that shrinks the coefficients towards zero. Consequently, coefficients with low explanatory power are “penalised” and thus forced out of the model (James *et al.*, 2013). This procedure is particularly effective for datasets with a large number of explanatory variables, as only coefficients that differ from zero and contribute substantial predictive power are included in the model as well (Storm *et al.*, 2019). The best-known regularisation methods are *LASSO* and *Ridge Regression*. Their optimisation functions are represented by minimising the sum of squares of risks (RSS)

² No fixed effects model is estimated as this could correct the model for an individual effect of farms. This is not necessary here, since the paper is primarily investigating predictive accuracy rather than causality.

in (7) and an additional penalty term, as in equations 8 (*LASSO*) and 9 (*Ridge*), according to James *et al.* (2013). Here, the coefficients $\hat{\beta}_j$ are minimised:

$$RSS = \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 \quad (7)$$

$$\min_{\beta_j} (RSS + \lambda \sum_{j=1}^p |\beta_j|) \quad (8)$$

$$\min_{\beta_j} (RSS + \lambda \sum_{j=1}^p \beta_j^2) \quad (9)$$

Both regularisation methods use a tuning parameter λ , which can be determined by the data using cross-validation. The influence of the regularisation increases with increasing λ (James *et al.*, 2013). The methods differ in the configuration of their penalty terms, also called L_1 - (*LASSO*) and L_2 -term (*Ridge*). The elastic net (EN) penalty term is shown in equation 10 and forms a combination of both penalty terms with $\alpha \in [0,1]$ (Zou und Hastie, 2005).

$$(1 - \alpha)\lambda \sum_j |\beta_j| + \alpha\lambda \sum_j \beta_j^2 \quad (10)$$

The EN uses automatic variable selection via the L_1 -part, which improves the prediction by inducing a grouping effect during variable selection so that highly correlated features are reduced (Ogutur *et al.*, 2011). In the present case, $\alpha = 0.5$ is assumed and the tuning parameter λ is validated with a 10-fold cross-validation.

4.2.2. Decision trees

Decision trees are used to divide a dataset into maximally homogeneous subgroups based on existing features (Storm *et al.*, 2019). Thus, a so-called “Classification and Regression Tree” (CART) creates binary splits in each case (Breiman *et al.*, 1984). For the Random Forest algorithm, the dataset is randomly split into subgroups of observations and features, with each subgroup differing only slightly. For each subgroup, a decision tree is created, from which the algorithm forms an average and, based on that, a prediction (Breiman, 2001). Another way to optimise the predictive power of decision trees is the method of boosting. This involves using

the residuals of the previous models so that the subsequent tree can specialise in them (Schapire, 1990). For the prediction of milk production here, two boosting algorithms are applied: Xtreme Gradient Boosting (XGB) and Gradient Boosting Machine (GBM). For both, the following hyperparameters are selected: number of trees = 2000; learning rate = 0.01; maximum depth = 5; proportion subgroup = 0.65; 10-fold cross-validation.

4.2.3. Neural networks

Typical neural networks (NNs) or multilayer perceptrons (MLPs) are structured as shown in Figure 3. They consist of several connected layers, each with multiple neurons. Their goal is to approximate a function f that learns from input values x to define an output vector y . NNs are basically a mapping of $y = f(x; w)$, where the weights w are continuously updated until they generate the best approximation of the function (Goodfellow *et al.*, 2016). The weights and neurons in the centre of the network cannot be observed directly, so they are called the hidden layer (Hastie *et al.*, 2017).

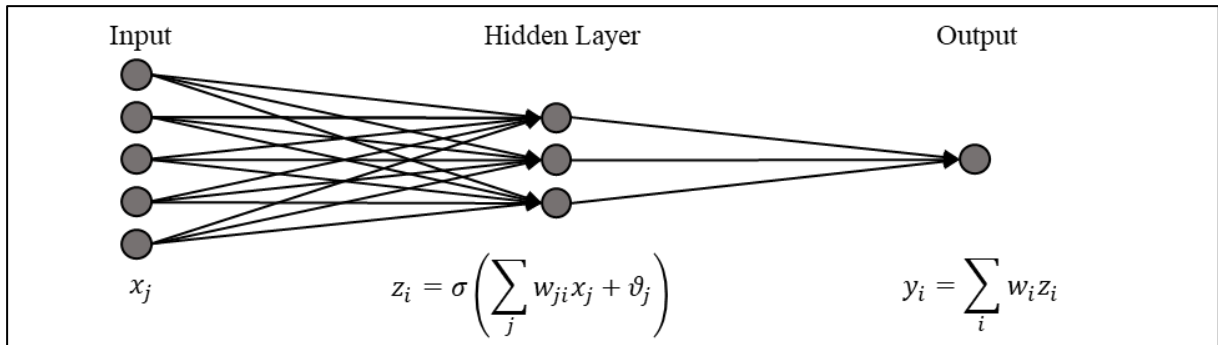


Figure 3. Schematic structure of a neural network

The strength of the connection between neurons is represented by the weights, which are multiplied by the input values and added to a constant ϑ that acts as the initial value for the optimisation (Athey and Imbens, 2019). The sum of these products is fed to the hidden layer via the activation function σ of the rectified linear units (ReLU) type (see Equation 11 and Fig. 3).

$$\sigma(z) = \max\{0, z\} \quad (11)$$

In the specific case of Figure 3, the activated values are then fed into the output layer, although this serves only as a simplification. NNs are called nets because of their interconnection of different functions. In this case, three hidden layers are used so that a chain structure in the form of $f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x)))$ occurs. The weights are updated and optimised by a loss function, similar to the optimisation of a linear regression (Goodfellow *et al.*, 2016). The following hyperparameters are chosen for the application of the NN for milk quantity forecasting: three hidden layers with 200, 100 and 50 neurons; optimisation with backpropagation; 150 epochs; stopping criterion = RMSE; minimum batch size = 32; 10-fold cross-validation.

4.2.4. Stacking

As suggested by its name, the stacking method combines several algorithms via a so-called “super-learner”. The aim is to improve the prediction by the best possible combination of different models. The learning algorithm can be chosen arbitrarily, but the application of the super-learner requires each of the previous models to be trained with the same V -fold cross-validation. The resulting predicted values of the base learner L are now combined into a matrix with dimension $V \times L$, which serves as the data basis for the super-learner (van der Laan *et al.*, 2007). For this reason, the stacking method is also referred to as a refined version of cross-validation that improves the winner-takes-all principle of cross-validation (Wolpert, 1992). In this case, a random forest algorithm is used to combine the EN, GBM and NN. The XGB model is not included in the stacking model here because the learning is very similar to that of the GBM and hence would be of limited added value.

5. Results and discussion

To evaluate the different algorithms and models, the trained algorithms are applied to the test set. Based on the most recent observations from 01/2018 to 09/2019, the monthly milk

production for all farms in this time horizon is predicted for the test dataset. Through this approach, different values for the relative and absolute measure of accuracy of the models can be observed. Therefore, the coefficient of determination (R^2) is compared to measure the relative error (Fig. 4) and the root mean square error (“RMSE”) for an absolute error value (Fig. 5). The adjusted R^2 does not need to be calculated in this instance because the explanatory power of specific variables is not primarily of interest; instead the focus is on the overall forecast accuracy, without giving consideration to degrees of freedom.

Figure 4 shows the differences of R^2 in the comparison between the algorithms, but also the influence of the feature selection by the difference between the models. The names of the models correspond to the abbreviations given in Table 1. It is noticeable that both linear approaches of OLS and EN show a significantly lower coefficient of determination (maximum 0.77 for OLS in model A) compared with the non-linear approaches. This difference is particularly prominent in Model E, where all variables are included. With RMSE of about 3.2 %, the linear model has no explanatory power for predicting milk production. This indicates that the model is overfitting due to an excessive number of variables, and thus cannot predict or can only predict poorly with out-of-sample data from the test set. Preventing this is the task of the EN, which uses a penalty term to give less weight to less relevant variables. The results in Figure 4 suggest that this penalty term is heavily weighted in models A to E, resulting in underfitting. Comparing the non-linear approaches, there are only slight differences in the R^2 for the respective algorithms and between models A to E. With the exception of model E, algorithm GBM has the highest coefficient of determination (maximum 0.92 in model C). Comparing this with the schematically similar algorithm XGB, only small differences of up to 3 % R^2 can be seen. This can be explained by the same hyperparameter selection and the similar learning mechanism. The neural network has the lowest coefficient of determination of the non-linear algorithms in all models. The combination of the models as a stacking algorithm does show any significant improvement with regard to the R^2 of the other algorithms.

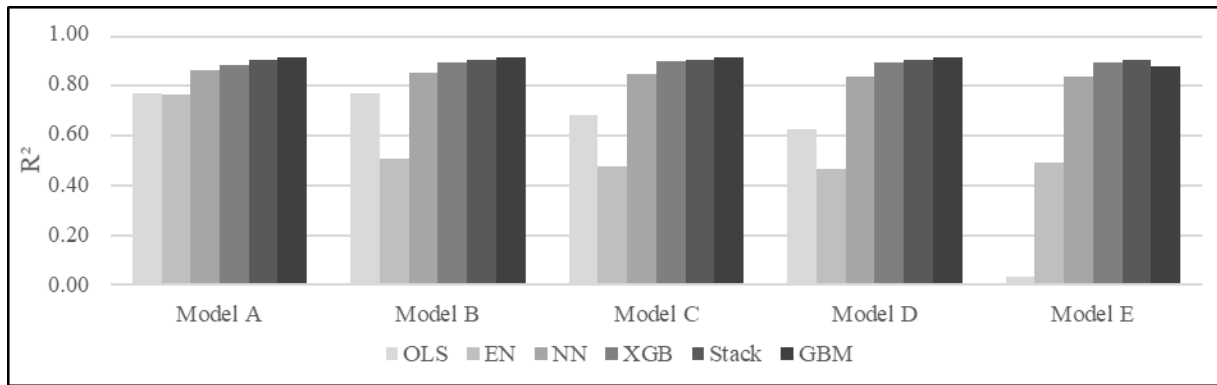


Figure 4. Coefficient of determination for all applied algorithms and models

The RMSE results in Figure 5 are similar to the R^2 shown in Figure 4. Figure 5 highlights that there is also a difference between linear and non-linear approaches in the observed RMSE. However, unlike the relative measure of the coefficient of determination, here the RMSE of the OLS and the EN are closer together. Looking at the value of the linear regression for model E, the problem of overfitting can also be observed in Figure 5. Figure 5 confirms that there is only a slight difference between the non-linear algorithms. Here, the RMSE of the NN stands out since the distance to the other models seems to be larger in relative terms here than in the R^2 comparison and is approximately at the mean between the values of the linear and non-linear models.

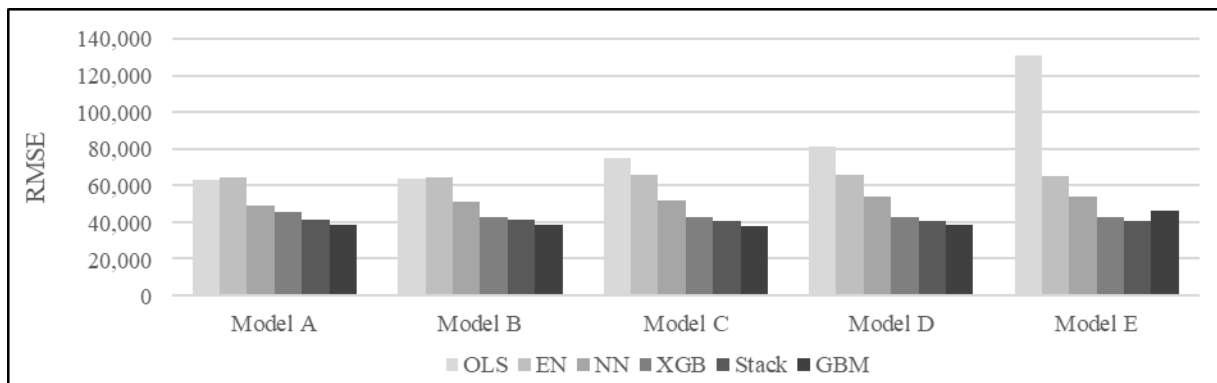


Figure 5. RMSE for all applied algorithms and models

One explanation for the difference in the results for R^2 and RMSE in linear and non-linear approaches could be the splitting of the dataset by the indicator function into moderately and rapidly growing farms. Figure 6 depicts the coefficient of determination for the respective groups for models A and C as examples to consider this relationship. The columns grouped side

by side show that in all combinations of model and algorithm, the R^2 of the group of fast-growing farms (group a) is lower in the linear approaches (OLS, EN). In particular, compared with the non-linear approaches, the R^2 in these respective groups is significantly lower, with a maximum value of 0.68 (OLS a, model A). This could be caused by investments as a driving factor of growth in group a. Milk production in this group does not increase linearly, but partly in step changes due to investments in new barn constructions and/or expansion. This development can be explained more effectively by a non-linear algorithm. Thus, these algorithms show an opposite distribution in that the group of rapidly growing farms can be predicted better than the other group. The minimum coefficient of determination of “group b” is 0.77 (XGB b, model A) and thus above the maximum value from the comparison group of linear models.

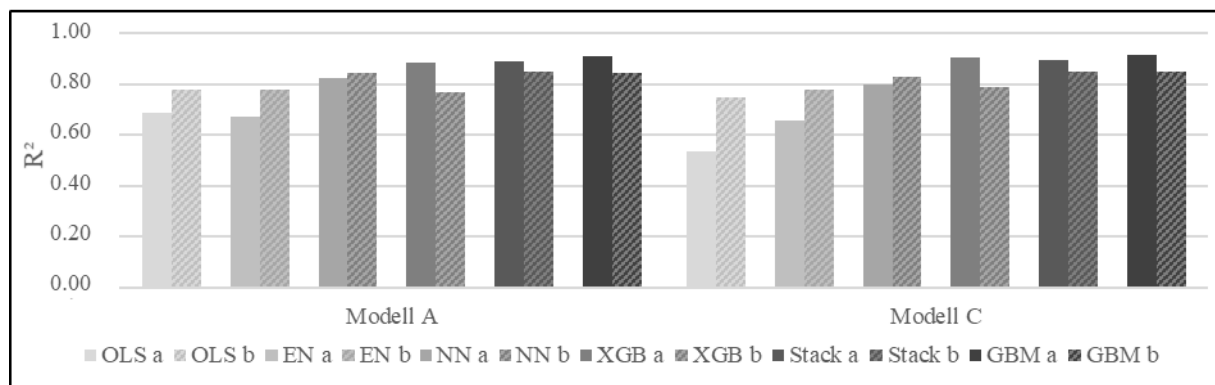


Figure 6. Coefficient of determination in the comparison of farm groups a and b

Summing observed and predicted values of milk production allows a graphical view to be given of the actual and predicted milk quantity over time for the dataset. Figure 7 shows the monthly milk quantity in ECM t for the period 01/2018 to 09/2019. It should be noted that only 11 data points are available for each year and farm. Significant drops in milk volume, such as in 07/2018, can also be related to this data distribution.

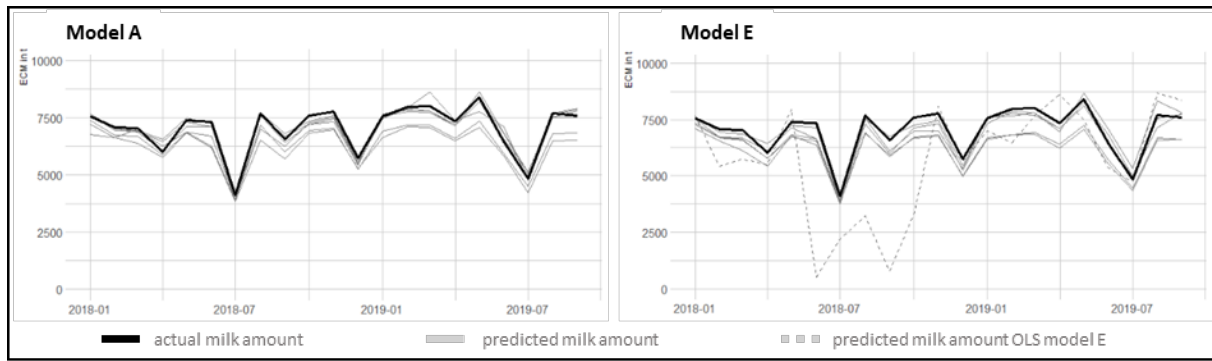


Figure 7. Predicted values in total in the test period (01/18 – 09/19)

The graph above shows that the predicted values of model A and E differ only slightly in their sum. The actual observed milk volume is shown as a black line, while the prediction of the models is shown in grey. Again, the differences within the models between the algorithms are marginal. In contrast, the shape of the OLS regression in model E (dotted line) is remarkable. As already evident in the results of R^2 and RMSE, the model is overfitted by too many variables, and thus becomes useless for prediction purposes. It is interesting to note that all the models tend to underestimate the amount of milk. However, overall it can be stated that all the applied models roughly represent the trend of milk production.

Another approach to compare the models focuses on the time required for training the algorithms. This is especially interesting for potentially extending the dataset by a larger number of farms in future. The differences between the non-linear approaches can be seen in a comparison of the training time of the models in Table 2. Here, the XGB algorithm, with an average training time of about 47.5 seconds for the respective models, is clearly ahead of the NN, GBM and stacking method, which require about 30 to 55 times the training time on average. Stacking also takes longer than the other methods since it adds the training times of the combined methods and its own execution time. Meanwhile, the training times of the linear models are considerably faster at about 4.6 seconds (EN) and 0.5 seconds (OLS).

Table 2
Training time of the algorithms in seconds

	OLS	EN	NN	XGB	Stack	GBM
Model A	0.05	0.43	387.30	12.70	571.74	184.01
Model B	0.05	1.63	405.00	16.15	606.74	200.11
Model C	0.20	1.94	840.42	27.86	1814.90	972.54
Model D	0.28	2.30	934.50	32.68	2037.62	1100.82
Model E	0.92	7.52	1263.12	53.07	3072.74	1802.10
Mean	0.50	4.61	1276.78	47.49	2701.25	1419.86

In addition to the prediction accuracy and speed of the models, it is also of interest to identify which variables have the greatest relative influence on the prediction of milk quantity production. There are different methods for measuring the importance of each variable. In this case, a standardised variable importance procedure using “partial dependence plot” (PDP) according to Greenwell *et al.* (2018) is applied. Representing the estimation methods used, OLS, XGB, and NN are compared. When calculating variable importance for OLS, the values correspond to the absolute values of the t-statistic. For XGB, the percentage increase in MSE is calculated when a particular variable is included in the overall model. The variable importance of a neural network is calculated based on the weights of the connected neurons. Figure 8 presents a comparison of these three algorithms for model A. For clarity, only the top 15 variables are listed by importance, and the importance of the dummies per farm is excluded. The absolute values of the respective algorithms cannot be compared; instead their relative distribution is considered. First it is noticeable that for all algorithms one of the time variables is significant. Especially for the XGB model, the variable “date” has a significant influence on the prediction compared with the other variables. This corresponds to the time trend by the combination of the variables “year” and “month”. For the OLS regression, the variables “year” and the different months are of greater importance. The variable “year” is also highly relevant

for the NN for the forecast, and some weather variables such as “wind” and “precipitation” are essential for the prediction. Thus, the NN differs somewhat from the OLS in its order of variable importance and from the XGB in its relative relationship. In all three cases, the milk and feed price variables are of relatively little importance as predictors within the models.

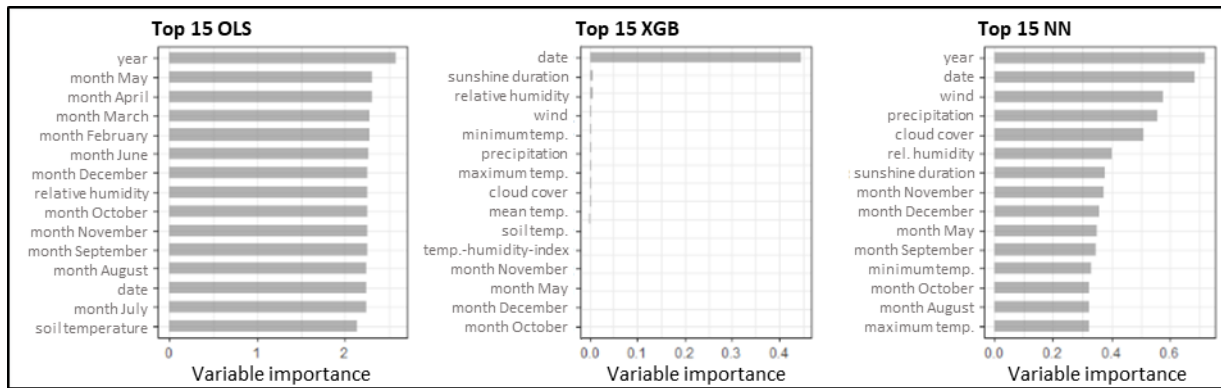


Figure 8. Variable importance for prediction in model A

However, when considering Figure 8, it should be noted that the comparison of feature importance is not equivalent to the statistical significance found in a linear regression model. While the values listed in Figure 8 are important for prediction, they are not a precise measure of the variable’s influence on the results. For example, a variable may be of low importance for prediction due to a high correlation with another variable (e.g. minimum and maximum temperature), but still be a weighted as a determinant for the dependent variable (Ifft *et al.*, 2018).

6. Conclusions

The comparison of different methods of ML with each other and with OLS regressions provides information about which of these methods may be particularly suitable for forecasting regional milk production. Four non-linear approaches (XGB, GBM, NN, stacking) and two linear approaches (OLS, EN) are compared. The difference between the two approaches is evident in both relative and absolute indicators of model selection: the non-linear algorithms considered each have higher coefficients of determination and lower RMSE. The specification of the

models with respect to variable and lag structure is less important for prediction quality than the choice of the estimation approach. The comparison of the importance of the variables to the prediction of the models shows that the small differences between the models A to E are due to the greater importance of time and trend variables compared with weather variables. These include unrecorded weather effects and management decisions, which cannot be directly accounted for with the available data. The time trend can be explained better by non-linear approaches, especially for the group of more rapidly growing farms. The comparison of non-linear algorithms with respect to the accuracy of their prediction only shows a difference with respect to the time required to train the models. A balance needs to be struck here between greater accuracy (GBM) and faster computational performance (XGB). This decision is highly relevant, especially for forecasts based on large datasets. In conclusion, an appropriately large dataset in connection with the algorithms mentioned above allows a geographical analysis to be undertaken for the entire federal state of Lower Saxony, meaning that forecasts can be produced not only for the milk quantity of individual farms, but for individual municipalities and districts as well. For this geographical approach, the indicator function used could easily consider further effects of regional structural change.

References

- Adediran, S., Ratkowsky, D., Donaghy and D., A. Malau-Aduli (2012): Comparative evaluation of a new lactation curve model for pasture-based Holstein-Friesian dairy cows. *Journal of Dairy Science*, Vol. 95, pp. 5344-5356.
- Athey, S. and G.W. Imbens (2019): Machine Learning Methods Economists Should Know About. *Annual Review of Economics* 2019, Vol. 11, pp. 685-725.
- Baudracco, J., Lopez-Villalobos, N., Holmes, C., Comeron, E., Macdonald, K., Barry, T. and N. Friggens (2011): e-Cow: an animal model that predicts herbage intake, milk yield and

- live weight change in dairy cows grazing temperate pastures, with and without supplementary feeding. *Animal*, Vol. 6, pp. 980-993.
- Blayney, D. and R. Mittelhammer (1990): Decomposition of Milk Supply Response into Technology and Price-Induced Effects. *American Journal of Agricultural Economics*, Vol. 72, pp. 864-872.
- BLE – Bundesanstalt für Landwirtschaft und Ernährung (2020a): Erzeugung von Milch, MilCHFett und Milcheiweiß in Deutschland nach Kalenderjahren, URL: <https://www.ble.de/DE/BZL/Daten-Berichte/Milch-Milcherzeugnisse/Mengen.html?nn=8906974>.
- BLE – Bundesanstalt für Landwirtschaft und Ernährung (2020b): Bericht zu Markt- und Versorgungslage mit Milch und Milcherzeugnissen. Berlin.
- BMEL – Bundesministerium für Ernährung und Landwirtschaft (2020a): Produktionswert des Bereichs Landwirtschaft, landwirtschaftliche Gesamtrechnung, URL: <https://bmel-statistik.de/landwirtschaft/landwirtschaftliche-gesamtrechnung/>
- BMEL – Bundesministerium für Ernährung und Landwirtschaft (2020b): Preise für konventionell erzeugte Kuhmilch. Verschiedene statistische Monatsbericht, 2010-2020.
- Bouamra-Mechemache, Z., Jongeneel, R. and V. Requillart (2008): Impact of a gradual increase in milk quotas on the EU dairy sector. *European review of agricultural economics*, Vol. 35, pp. 461-491.
- Breiman, L. (2001): Random Forests. *Machine Learning*, Vol. 45, pp. 5–32
- Breiman, L., Friedman, J.H., Olshen, R.A. and C.J. Stone (1984): Classification and Regression Trees. Wadsworth and Brooks, Monterey, CA.
- Brügemann, K., Gernand, U., von Borstel, U. and S. König (2011): Genetic analyses of protein yield in dairy cows applying random regression models with time-dependent and

- temperature x humidity-dependent covariates. *Journal of Dairy Science*, Vol. 94, pp. 4129-4139.
- Collard, B., Boettcher, P., Dekkers, J., Petitclerc and L. Schaeffer (2000): Relationships Between Energy Balance and Health Traits of Dairy Cattle in Early Lactation. *Journal of Dairy Science*, Vol. 83, pp. 2683-2690.
- FAOSTAT (2020): Statistical database of the Food and Agriculture Organization of the United Nations. Rome, URL: <http://www.fao.org/faostat/en/#home>.
- Goodfellow, I., Bengio, Y. and A. Courville (2016): Deep Learning. MIT Press, Cambridge, MA.
- Greenwell, B., Boehmke, B. and A. McCarthy (2018): A simple and effective model-based variable importance measure. arXiv preprint:1805.04755.
- Hallam, A. (1991): Economics of Size and Scale in Agriculture: An Interpretive Review of Empirical Measurement. *Review of Agricultural Economics*, Vol. 13, pp. 155-172.
- Harder, I., Stamer, E., Junge, W. and G. Thaller (2019): Lactation curves and model evaluation for feed intake and energy balance in dairy cows. *Journal of Dairy Science*, Vol. 102, pp. 7204-7216.
- Hastie, T., Tibshirani, R. and J. Friedman (2017): The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, New York, NY.
- Hill, D. and E. Wall (2017): Weather influences feed intake and feed efficiency in a temperate climate. *Journal of Dairy Science*, Vol. 100, pp. 2240-2257.
- Hoehl, S. and S. Hess (2021): Liquidity, hedging and the survival of North German dairy farms. *European Review of Agricultural Economics*, Vol. 00, pp. 1-28.
- Ifft, J., Kuhns, R. and K. Patrick (2018): Can machine learning improve prediction – an application with farm survey data. *International Food and Agribusiness Management Review*, Vol. 21, pp. 1083-1098.

- James, G., Witten, D., Hastie, T. and R. Tibshirani (2017): An Introduction to Statistical Learning. Springer, New York, NY.
- Kolver, E. and L. Muller (1998): Performance and Nutrient Intake of High Producing Holstein Cows Consuming Pasture or a Total Mixed Ration. *Journal of Dairy Science*, Vol. 81, pp. 1403-1411.
- Larcher, M., Schönhart, M., Schmid, E. and S. Vogel (2015): Intensivierungspläne österreichischer MilchproduzentInnen angesichts der auslaufenden Milchquotenregelung 2015. *German Journal of Agricultural Economics*, Vol. 64., pp. 148-168.
- LSN – Landesamt für Statistik Niedersachsen (2020): Statistische Berichte Niedersachsen, CIII 1 – hj 1 / 2020.
- Mader, T., Davis, M. and T. Brown-Brandl (2006): Environmental Factors Influencing Heat Stress in Feedlot Cattle. *Journal of Animal Science*, Vol. 84, pp. 712-719.
- Mullainathan, S. and J. Spiess (2017): Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, Vol. 21, pp. 87-106.
- Neuenfeldt, S., Gocht, A., Heckeley, T. and P. Ciaian (2019): Explaining farm structural change in European agriculture: a novel analytical framework. *European Review of Agricultural Economics*, Vol. 46, pp. 713-768.
- OGUTU, J.O., SCHULZ-STREECK, T. and H.P. PIEPHO (2012): Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. *BMC Proceedings*, Vol. 6, pp. 10-16.
- Parsons, D., Rey, D., Tanguy, M. and I. Holman (2019): Regional variations in the link between drought indices and reported agricultural impacts of drought. *Agricultural Systems*, Vol. 173, pp. 119-129.
- Schapire, R. (1990): The Strength of Weak Learnability. *Machine Learning*, Vol. 5, pp. 197-227

- Schlenker, W. and M. Roberts (2009): Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 106, pp. 15594-15598.
- Sjaunja, L.O., Baevre, L., Junkkarinen, L. and J. Pedersen (1990): A Nordic proposal for an energy corrected milk (ECM) formula. 27th Session of ICRPMA, pp. 156-157.
- Storm, H., Baylis, K. and T. Heckelei (2019): Machine learning in agricultural and applied economics. *European Review of Agricultural Economics*, Vol. 47, pp. 1-44.
- Tekerli, M., Akinci, Z., Dogan, I. and A. Akcan (2000): Factors Affecting the Shape of Lactation Curves of Holstein Cows from Balikesir Province of Turkey. *Journal of Dairy Science*, Vol. 83, pp. 1381-1386.
- Van der Laan, M., Polley, E. and A. Hubbard (2007): Super Learner. *Statistical Applications in Genetics and Molecular Biology* Vol. 6.
- Vitali, A., Felici, A., Esposito, S., Bernabucci, U., Bertocchi, L., Maresca, C., Nardone, A. and N. Lacetera (2015): The effect of heat waves on dairy cow mortality. *Journal of Dairy Sciences*, Vol. 98, pp. 4572-4579.
- Wolpert, D. (1992): Stacked generalization. *Neural networks*, Vol. 5, pp. 241-259.
- Wood, P. (1967): Algebraic model of the lactation curve in cattle. *Nature*, Vol. 216, pp. 164-165.
- Wood, P. (1969): Factors affecting the shape of the lactation curve in cattle. *Animal Production*, Vol. 11, pp. 307-316.
- Zou, H. and T. Hastie (2005): Regularization and variable selection via the elastic net. *Journal of the royal statistical society*, Vol. 67, pp. 301-320.