Prediction of litter performance in lactating sows using machine learning, for precision livestock farming

Raphaël Gauthier, Christine Largouët, Jean-Yves Dourmad

Prediction of litter performance in lactating sows using machine learning, for precision livestock farming

Raphaël Gauthier[a,b,*], Christine Largouët[c], Jean-Yves Dourmad[a]

[a] *PEGASE, INRAE, Institut Agro, 35590, Saint Gilles, France*
[b] *Univ Rennes, CNRS, Inria, IRISA - UMR 6074, F-35000 Rennes*
[c] *AGROCAMPUS OUEST / INRIA, Univ Rennes, CNRS, IRISA, F-35000 Rennes*

Abstract

Predicting litter performance in lactating sows is an essential step towards the development of decision support systems for precision feeding in lactating sows. Numerous factors affecting litter performance have been described in literature. However, predictive models working on-farm in real time are not available. The main objectives of this research was to (i) explore 4 different machine learning strategies, and (ii) identify the best supervised learning algorithm in order to obtain reliable predictions of litter performance. This study was carried out with data obtained from 6 experimental farms over the last 20 years. Algorithms were trained to predict the litter weight at weaning using a set of 4 numeric and 3 categorical features, and a method for predicting secondary litter performance and nutrient output in milk from the predicted litter weight at weaning was evaluated. To evaluate the reliability of predictions within each farm, the mean error per farm (ME_f) and the mean absolute percentage error per farm ($MAPE_f$) were computed. The best performance for the prediction of litter weight at

Abbreviations: CD, critical difference; PLF, precision livestock farming; PF, precision feeding; R^2, coefficient of determination; RSD, residual standard deviation; LR, Linear Regression; LASSO, Least Absolute Shrinkage and Selection Operator; kNN, k-nearest-neighbors; RF, Random Forest; SVR, Support Vector for Regression; GTB, Gradient Tree Boosting; VR, Voting Regressor; MLP, Multi Layer Perceptron; AllFarms, all farms learning strategy; NoFarmFeature, no farm feature learning strategy; LOFO, leave-one-farm-out learning strategy; PerFarm, per farm learning strategy.

*Corresponding author
Email address: raphael.gauthier@gmx.com (Raphaël Gauthier)

weaning was obtained with an ensemble algorithm with farm-level training and testing (ME_f = -0.14 kg; $MAPE_f$ = 9.01%), but performance with simple linear regression was very close ($MAPE_f$ = 9.30%). Learning across all farms only achieved comparable results with the neural networks algorithm, but at higher computational costs. The method for predicting secondary litter performance and nutrient output from the predictions of litter weight at weaning reveals that the ME_f remains close to 0, and that the $MAPE_f$ only increases by a few percentage points. This study confirms the effect of numerous factors known in the literature to affect litter performance, such as litter size and parity of sows, but also revealed huge variations between farms. According to this study, reliable predictions could be obtained with interpretable supervised algorithms trained at farm level, with features that can be easily measured on-farm. This study thus shows that on-farm data are necessary to accurately train models and make reliable predictions at farm level. These predictions could be used by decision support systems in order to develop precision feeding approaches in lactating sows.

Keywords: Litter performance, Lactating sows, Supervised learning, Precision livestock farming

1. Introduction

Precision livestock farming (PLF) is a novel approach in livestock production systems that relies on intensive use of technology and process engineering to improve livestock sustainability and efficiency (Wathes et al., 2008; Pomar et al., 2019). As part of PLF, precision feeding (PF) principles are based on on-line measuring devices, computational methods, and feeding devices that make it possible to feed animals individually, with the right amount of nutrients provided at the right time (Pomar et al., 2019; Gaillard et al., 2020). The variability of nutrient requirements according to physiological stage has been successfully integrated into traditional nutrition models in swine such as InraPorc® (Dourmad et al., 2008) and NRC (2012). However, handling the variability of nutrient requirements over time and between animals is a major area of study that PF systems should take into account in order to improve the overall efficiency of the livestock feeding chain (Vranken and Berckmans, 2017; Pomar et al., 2019).

In lactating sows, milk production leads to large, variable nutrient requirements among individuals (Noblet et al., 1990; Gauthier et al., 2019). In contrast with dairy cows, available methods to directly measure sow milk production are either unaffordable and labour intensive at farm level, or may reduce piglet growth (Quesnel et al., 2015). However, measurements of litter performance can be used as proxies for milk production and provide estimates of nutrient output in milk at lower expense (Noblet et al., 1990; Hansen et al., 2012; Quesnel et al., 2015). The effects of numerous factors affecting litter performance have been reported in the literature. Litter size is considered to be the main factor influencing milk production and litter performance because it affects the number of functional mammary glands (Auldist et al., 1998; Ngo et al., 2012a), followed by stage of lactation, parity of sows, nutrition, and environmental factors such as temperature (Quesnel et al., 2015). Furthermore, genetic selection in maternal lines over the past decades has resulted in a dramatic increase in sow prolificacy and milk production combined with increased variability in litter performance (Silalahi et al., 2016).

Numerous factors affecting litter performance have been described in literature. However, predictive models working on-farm in real time are not available, although this is precisely what PF systems require for lactating sows (Gauthier et al., 2019). There is therefore a need for processes providing reliable predictions of litter performance from simple measurements. The main objective of this research was to explore different learning strategies along with different supervised machine learning algorithms to obtain reliable on-farm predictions of litter weight at weaning (LWW), as a proxy for milk production. This study was carried out with data obtained in 6 experimental farms over the last 20 years.

2. Material and Methods

2.1. General Outline

The questions raised by the prediction of LWW in lactating sows, in the context of precision livestock farming, included (i) Which machine learning algorithm is best adapted for this data? and (ii) Which is the best learning strategy to reliably use the resulting machine learning system on-farm? Litter weight at weaning is an easy-to-measure phenotype that is closely related to milk production in lactating sows. Assuming that piglets were mainly fed with milk, it was thus selected as the target of machine learning algorithms. In this study, real observations of LWW were measured on-farm by farmers or using automatic connected scales, thus, predicting LWW comes under the field of supervised machine learning. Moreover, because the outcome to be predicted is a continuous value, different supervised learning algorithms dedicated to regression tasks were selected. In order to identify the best way to generalize the machine learning algorithms on new data for a given farm, several learning strategies, each being a different way of splitting the database, were defined (section 2.3, Figure 1) and algorithms were trained according to a 5-fold cross-validation scheme (section 2.4). In the following sections, we describe the database, the different supervised algorithms that were used, and the different learning strategies evaluated for predicting LWW. After that, we evaluate the prediction of other litter performance drawn from the predicted LWW and used for the calculation of nutrient requirements (section 2.6).

2.2. Data preparation and preliminary statistical analysis

A database with 23,259 observations from 6 different farms was used to train supervised learning algorithms. The data were collected between January 2000 and January 2019. The dataset was composed of 4 numeric features, namely the duration of the lactation, the litter size at birth (LSB), the litter size at weaning (LSW), and the litter weight at birth (LWB), and 3 categorical features, namely the parity of sows, the month of farrowing, and the farm name. Cleaning steps of the database have been performed to remove unreliable and uncommon observations from the database. Observations presenting a negative LWW or greater than 300 kg, a lactation period shorter than 20 days or longer than 35 days, a LWB higher than 80 kg, or a parity greater than 20 were removed. Observations with one or several missing features were also removed. The data set was thus reduced to 20,368 complete observations.

A preliminary statistical analysis of the database was conducted on the features, the LWW target, and on other performance criteria that are commonly used to evaluate litter performance and calculate nutrient requirements for milk production in lactating sows. The litter weight gain (LWG, in kg) was computed as the difference between LWW and the total weight at birth of weaned piglets. The litter average daily gain (LADG, in kg/d) was computed as LWG divided by the duration of lactation. The piglet weight at birth (PWB, in kg) was obtained by dividing LWB by LSB. The piglet weight at weaning (PWW, in kg) was obtained by dividing LWW by LSW. The piglet weight gain (PWG, in kg) was computed as the difference between LWW and PWB. The piglet average weight gain (PADG, in g/d) was computed as LADG divided by the LSW. The dry matter ($\mathrm{DM_m}$), energy in milk ($\mathrm{E_m}$), and nitrogen in milk ($\mathrm{N_m}$), were computed according to the equation in Noblet and Etienne (1989):

$$DM_m, kg/d = (0.72 \times LADG - 7 \times LSW) \div 1000 \tag{1}$$

$$E_m, MJ/d = (20.6 \times LADG - 376 \times LSW) \div 1000 \tag{2}$$

$$N_m, g/d = 0.0257 \times LADG + 0.42 \times LSW \tag{3}$$

Statistical analyses of the database was performed with Python 3 using the ANOVA linear model (statsmodels 0.11.1 Seabold and Perktold, 2010), with statistical significance of $P < 0.05$. Parity was analyzed for the fixed effects of farm and month of farrowing. The duration of lactation, LSB, LWB, PWB and LSW were analyzed for the fixed effects of farm, parity and month of farrowing. PWW, PADG, LWW, LADG were analyzed for the fixed effects of farm, parity, litter size at weaning, and month of farrowing.

2.3. Learning strategies

In order to identify the best strategy able to generalize the machine learning algorithms for a given farm, four learning strategies (Figure 1) were explored, in combination with the 8 supervised learning algorithms described in section 2.4. These learning strategies were chosen in order to evaluate the genericity of the adjusted algorithms, depending on the availability of data in the farm where the algorithm is to be used. In the AllFarms learning strategy, machine learning algorithms were trained on 80% of the database and tested on the remaining part, with all the features, including farm name feature. This strategy aims at discovering how algorithms could map the relation between target and features in a multi-domain context. Splitting procedure is stratified to ensure that farm frequencies are the same for training and testing. With this strategy, the obtained algorithms are common to the six farms and might be used for prediction in each of these farms, considering a farm effect. In the NoFarm-

Feature learning strategy, machine learning algorithms were trained and tested in the same way as for the AllFarms strategy, but with less supervision since the farm name feature was removed from the set of available features. With this strategy, prediction might be performed without any information on the farm, *i.e.*, for a new farm. In the leave-one-farm-out (LOFO) learning strategy, the splitting, training, and testing procedures were conducted recursively, *i.e.*, once per farm. Machine learning algorithms were trained with the full data sets provided by 5 farms, and tested on the full data set of the remaining farm, in order to check if it is possible to predict accurately data from a different domain than the one used for training an algorithm. With this strategy, the obtained algorithms might be used for prediction on a farm not considered at training (Bascol et al., 2017). In the PerFarm learning strategy, machine learning algorithms were trained recursively and tested farm by farm, with respect to an 80/20 ratio between training and testing sets. With this strategy, a farm-specific algorithm is adjusted and used on each farm, which requires having a sufficient amount of data from the farm for learning.

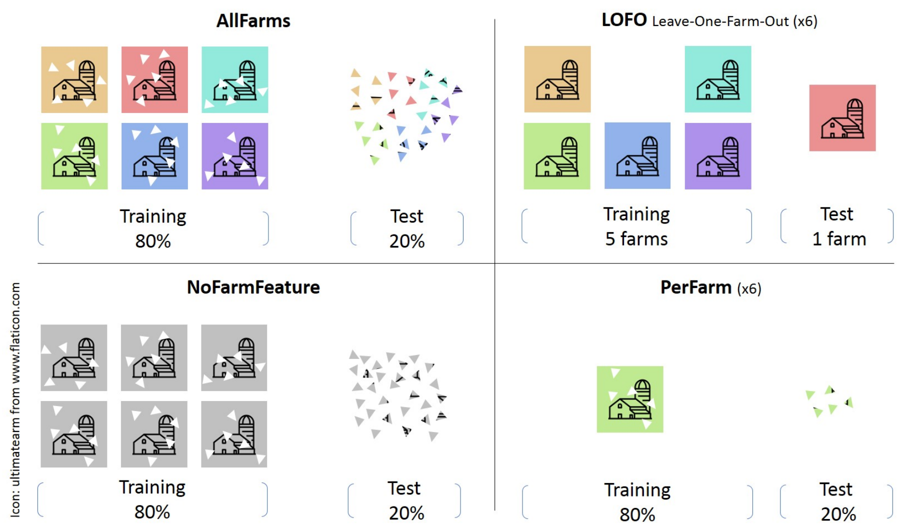

Figure 1: Training and test sets according to the different learning strategies. Color represents the farm name feature. Grey represents a learning strategy where the farm name feature is removed.

Prior to training, parity and month of farrowing features were encoded as one-hot vectors for all learning strategies. The farm name feature was transformed into one-hot vectors in the AllFarms and LOFO learning strategies.

2.4. Algorithm selection & hyperparameter tuning

Each learning strategy was tested with 8 supervised learning algorithms that can be used for regression tasks and are provided in the scikit-learn library (Pedregosa et al., 2011). These algorithms are the Linear Regression (LR), the Least Absolute Shrinkage and Selection Operator (LASSO), the k-Nearest-Neighbors (kNN), the Random Forest (RF), the Gradient Tree Boosting for regression (GTB), the Support Vectors for Regression (SVR), the Voting Regressor (VR) and the Multi-Layer Perceptron (MLP). The 8 supervised algorithms choosen for the experimentation are the more popular to achieve regression tasks (Géron, 2019) and are generally associated with high performance. As a preprocessing step, training and test data were standardized separately for LR, LASSO, kNN, GTB, SVR, VR, and MLP algorithms, regardless of the learning strategy.

Hyperparameter combinations were evaluated to identify the optimal training for each algorithm (Table 1). For any hyperparameter combination, algorithms were trained according to a 5-fold cross-validation scheme (Pedregosa et al., 2011). As no assumptions could be made on the right values of continuous hyperparameters, several power of tens were tried for the alpha hyperparameter of LASSO, for the C and epsilon hyperparameters of SVR, and for the alpha hyperparameter of MLP (Buitinck et al., 2013). For RF, a random search was first performed to reduce the hyperparameter space (Bergstra and Bengio, 2012). The average performance of an algorithm and a given combination of hyperparameters was computed from the performance obtained during the cross-validation.

Table 1: Hyperparameters and data transformation of the selected supervised learning algorithms[1]

Algorithm[2]	Preprocessing of features[3]	Hyperparameter name[4]	Hyperparameter values
LR	Standardization	fit_intercept	True, False
Lasso	Standardization	alpha	$[10^{-10}, 10^{1}] \cap \mathbb{Z}$
kNN	Standardization	n_neighbors	$[1, 50] \cap \mathbb{N}$
		weights	uniform, distance
RF[5]	-	max_depth	None
		min_samples_leaf	3,4,5
		min_samples_split	8, 10, 12
		n_estimators	300, 400, 500
GTB	Standardization	loss	ls, lad, huber
		n_estimators	1, 8, 16, 32, 64, $10^2, 2 \cdot 10^2, 5 \cdot 10^2, 10^3, 5 \cdot 10^3$
		learning_rate	$[0.01, 0.05, 0.1, 0.25, 0.5, 1]$
		max_depth	$[1, 8] \cap \mathbb{N}$
SVR	Standardization	kernel	rbf
		epsilon	0.1,1,10, 20, 100
		C	0.1,1,10, 20, 100
		gamma	scale, auto
VR	Standardization	estimators	LASSO, SVR, MLP
MLP	Standardization	hidden_layer_sizes	2, 10, 100, 200, 300, 400, 500
		activation	tanh, relu, logistic
		solver	lbfgs, sgd, adam
		alpha	$10^{-5}, 10^{-3}, 10^{-1}, 10^{1}, 10^{3}$

[1] Performance with one hyperparameter combination was evaluated 5 times according to a 5-fold cross-validation scheme.

[2] LR: Linear Regression, Least Absolute Shrinkage and Selection Operator: LASSO, kNN: k-nearest neighbors, RF: Random Forest, SVR: Support Vector for Regression, GTB: Gradient Tree Boosting, VR: Voting Regressor, MLP: Multi-Layer Perceptron

[3] Standardization of training and testing sets were carried out separately

[4] Hyperparameter names according to Buitinck et al. (2013)

[5] Hyperparameter space was first reduced with a random search on hyperparameters (Bergstra and Bengio, 2012)

2.5. Evaluation of learning strategies and algorithm performance

For a given learning strategy and a given algorithm, the hyperparameter combination that yields the best training performance according to the 5-fold cross-validation scheme was chosen to train the final model, and make predictions of LWW over the corresponding test set (Figure 1). The quality of predictions was assessed through two main criteria. The first was the mean error at farm level (ME_f):

$$ME_f = \frac{1}{n_f} \sum_{i=1}^{n_f} y_i - \hat{y}_i \tag{4}$$

where ME_f, $f \in [1,6] \cap \mathbb{N}$, is the ME of farm f in kilograms, n_f is the size of the test set in farm f, y_i is the i^{th} observation, and $\hat{y}_i$ is the i^{th} prediction.

The second was the mean absolute percentage error per farm ($MAPE_f$):

$$MAPE_f = \frac{1}{n_f} \sum_{i=1}^{n_f} \frac{\|y_i - \hat{y}_i\|}{y_i} \times 100 \tag{5}$$

where $MAPE_f$, $f \in [1,6] \cap \mathbb{N}$, is the MAPE of farm f as a percentage, n_f is the size of the test set in farm f, y_i is the i^{th} observation, and $\hat{y}_i$ is the i^{th} prediction.

A Nemenyi test was carried out to compare learning strategies based on the average ranks obtained by the 8 supervised algorithms across the 6 farms, according to the $MAPE_f$ metric (Demšar, 2006). The test was considered significant when $P < 0.05$.

2.6. Predictions of secondary litter performance criteria and nutrient output in milk

Further predictions of secondary litter performance criteria (LWG, LADG, PWW, PWG, PADG) and nutrient output in milk (DM_m, E_m, N_m) were computed. To obtain these secondary predictions, the transformations mentioned in section 2.2 were applied to predicted values of LWW obtained with the supervised algorithm and learning strategy that yielded the best performance. Predictions of LWG, LADG, PWW, PWG, PADG, DM_m, E_m, and N_m were then estimated at the farm level through the ME_f and $MAPE_f$ metrics.

3. Results

In this section, the results of the preliminary statistical analysis of the database are first presented. Then, learning strategies and algorithms are compared for the prediction of LWW, and predictions of secondary outcomes are evaluated.

3.1. Preliminary statistical analysis of the database

The results of the preliminary statistical analysis of the database are presented in table 2. The number of litters varied between farms from 272 in farm 5 up to 7,476 in farm 1. The effect of farm on parity was significant, with the lowest value of 1.9 (farm 5) and the highest value of 3.9 (farm 3; $P < 0.001$). However, the effect of farm did not explain much of the variability ($R^2 = 0.03$; $RSD = 1.9$). The effects of farm, parity, and month of farrowing were significant for the duration of lactation ($R^2 = 0.23$; $P < 0.001$). More precisely, weaned piglets were older in farms 1, 2, 3, and 6 (about 28 days) than in farms 4 and 5 (about 25 and 21 days, respectively). The effects of farm and parity were significant for LSB ($R^2 = 0.12$; $P < 0.001$). The smallest litter at birth was composed of 12.4 piglets (farm 1), and the largest was composed of 14.6 piglets (farm 3). The effect of farm, parity, and month of farrowing was significantly on LSW ($R^2 = 0.21$; $P < 0.001$). LSW differed greatly between farms, from 9.6 piglets (farm 1) up to 12.1 piglets (farm 3). The effects of farm, parity, and month of farrowing were also significant for PWB and LWB ($P < 0.001$). The effects of farm, parity, LSW and month of farrowing were significant on PWW, PADG, LWW, LADG, DM_m, E_m, and N_m ($P < 0.001$). The coefficient of determination was the lowest for PWB ($R^2 = 0.06$), intermediate for PWW, PADG and LWB (R^2 of 0.22, 0.18, and 0.19, respectively), and the highest for LWW, LADG, DM_m, E_m, and N_m (R^2 of 0.61, 0.54, 0.52, 0.50, and 0.57, respectively).

Table 2: Influence of farm, parity, litter size at weaning, and month of farrowing on lactation characteristics, litter performance, and nutrient output in milk[1]

	Farm						Statistics[2]					
	1	2	3	4	5	6	RSD[3]	R^2	F	P	LSW	M
Number of litters	7476	6467	2673	605	272	2875						
Parity	2.6	2.9	3.4	3.9	1.9	3.2	1.9	0.03	***	-	-	NS
Lactation length, d	27.9	27.9	28.7	25.3	21.4	27.9	1.8	0.23	***	***	-	***
LSB	12.4	13.3	14.6	13.8	13.8	13.5	3.4	0.06	***	***	-	NS
LSW	9.6	10.6	12.1	11.1	11.6	11.3	1.8	0.21	***	***	-	***
PWB, kg	1.41	1.43	1.45	1.54	1.3	1.51	0.28	0.06	***	***	-	***
PWW, kg	8.23	7.95	8.72	8	5.92	9.03	1.28	0.22	***	***	***	***
PADG, g/d	245	234	253	255	217	270	43	0.18	***	***	***	***
LWB, kg	17.1	18.6	20.8	20.4	17.1	20.1	4.7	0.13	***	***	-	***
LWW, kg	78.7	84	104.7	85.4	68.4	101.5	12.8	0.61	***	***	***	***
LADG, kg/d	2.34	2.46	3.03	2.7	2.5	3.03	0.42	0.54	***	***	***	***
DM_m, kg/d	1.62	1.7	2.1	1.87	1.72	2.1	0.3	0.52	***	***	***	***
E_m, MJ/d	44.6	46.8	58	51.5	47.2	58.1	8.7	0.5	***	***	***	***
N_m, g/d	64.2	67.8	83.1	74	69.2	82.6	10.8	0.57	***	***	***	***

[1] LSB: litter size at birth, LSW: litter size at weaning, PWB: piglet weight at birth, PWW: piglet weight at weaning, PADG: piglet average daily gain, LWB: litter weight at birth, LWW: litter weight at weaning, LADG: litter average daily gain, DM_m: dry matter in milk, E_m: net energy in milk, N_m: nitrogen in milk

[2] Data were analyzed using ANOVA linear models that included the effect of farm (F), parity (P), litter size at weaning (LSW), and month of farrowing (M). ***: $P < 0.001$; *: $P < 0.05$; NS: non significant; - : not included in the ANOVA.

[3] RSD: residual standard deviation

Mean LADG varied according to farm ($P < 0.001$), between 2.34 and 3.03 kg/d (Table 2, Figure 2a). Its variability was also dependent on the farm (Figure 2a), with a larger variability in farms 1, 2, and 6 than in farms 3, 4, and 5. Mean LADG increased according to LSW ($P < 0.001$), from 1.29 kg/d (LSW = 5) up to 3.20 kg/d (LSW = 15), whereas it slightly decreased for LSW above 15 (Figure 2b). PADG decreased when LSW increased, from 259 (± 77.5) g/d (LSW = 5) down to 159 (± 34.8) g/d (LSW = 17) (Figure 2c), with a concomitant decrease in its variability. With respect to parity, LADG increased ($P < 0.001$) from 2.43 (Parity 1) up to 2.68 kg/d (Parity 3) and slightly decreased for higher parities (Figure 2d). A similar trend was found for LWW. The month of the year affected ($P < 0.001$) LADG with the lowest values in July (2.51 kg/d) and the highest in December (2.65 kg/d).

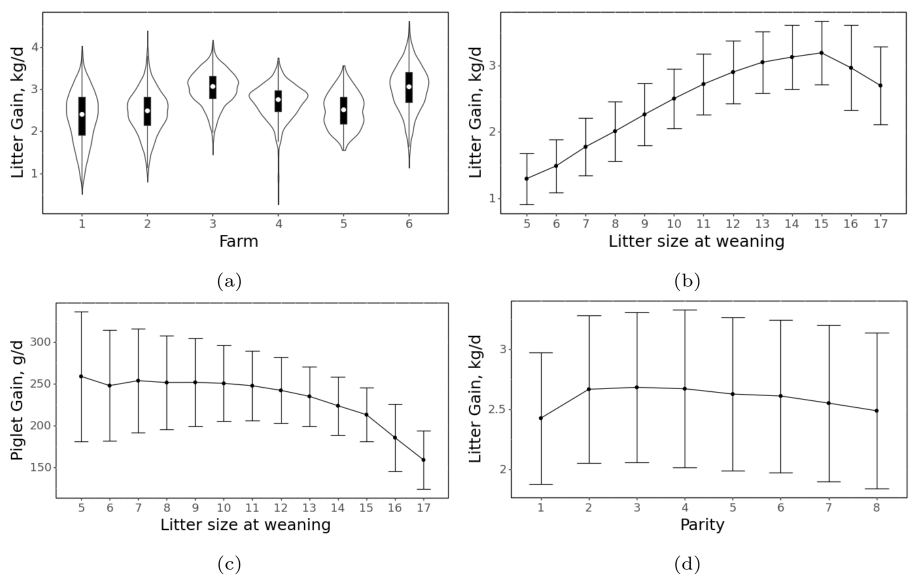

Figure 2: Influence of (a) farm on litter gain (kg/d), (b) litter size at weaning on litter gain, (c) litter size at weaning on piglet gain (g/d), and (d) parity of sows on litter gain. Violin plot with nested boxplot displays the data distribution vertically with a white point for the median value. Dots connected with a line represents mean values. ⊤ is equal to the mean plus one standard deviation. ⊥ is equal to the mean minus one standard deviation.

3.2. Comparisons between learning strategies and supervised learning algorithms

The quality of these predictions was first assessed at farm level according to the ME_f metric. These results are presented as box plots describing the farm dispersion of ME_f values (Figure 3), according to the learning strategy and the algorithm used. In the AllFarms learning strategy, the ME_f obtained for the prediction of LWW range from 0.0 ($\pm$0.29) kg (MLP) to 0.0 ($\pm$0.75) kg (LR), depending on the algorithm. In the NoFarmFeature learning strategy, the ME_f are higher than in the AllFarms strategy, with much more variability between farms. They vary between 0.0 ($\pm$5.59) kg (LR) and 1.0 ($\pm$4.27) kg (GTB), depending on the algorithm. The highest between farm variability in ME_f was obtained for the LOFO learning strategy with mean values ranging from -3.0 ($\pm$8.7) kg (Lasso) to 0.0 ($\pm$5.12) kg (MLP). Conversely, the lowest between farm variability in ME_f was obtained for the PerFarm learning strategy with values ranging from -1.0 ($\pm$0.24) kg (kNN) to 0.0 ($\pm$0.55) kg (MLP).

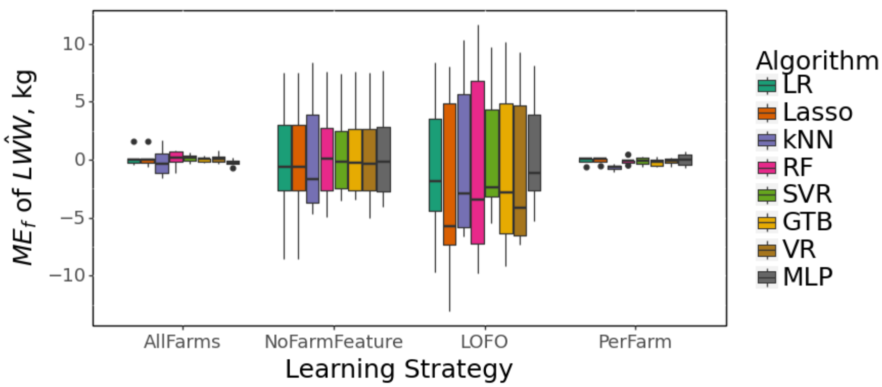

Figure 3: Box plots of mean error per farm (ME_f, in kg) for the predictions of litter weight at weaning (LWW) according to supervised algorithms and learning strategies. For each box plot, the black line represents the median value and whiskers represent 1.5 times the interquartile range (LR, Linear Regression; LASSO, Least Absolute Shrinkage and Selection Operator; kNN, k-nearest-neighbors; RF, Random Forest; SVR, Support Vector for Regression; GTB, Gradient Tree Boosting; VR, Voting Regressor; MLP, Multi Layer Perceptron; AllFarms, all farms learning strategy; NoFarmFeature, no farm feature learning strategy; LOFO, leave-one-farm-out learning strategy; PerFarm, per farm learning strategy)

The quality of the predictions of LWW assessed at farm level according to the $\mathrm{MAPE_f}$ metric are presented in figure 4 as box plots describing the farm dispersion of $\mathrm{MAPE_f}$ values, according to the learning strategy and the algorithm used. These values are expressed as a percentage of the mean value. The $\mathrm{MAPE_f}$ obtained for the prediction of LWW in the AllFarms learning strategy range from 9.0 (±2.06) % (MLP) to 11.0 (±2.98) % (Lasso) (Appendix A - Table A.5). The $\mathrm{MAPE_f}$ obtained in the NoFarmFeature learning strategy are higher than with the AllFarms strategy, with values ranging from 10.0 (±1.69) % (SVR) and 12.0 (±2.86) % (LR) (Appendix A - Table A.6). The $\mathrm{MAPE_f}$ obtained in the LOFO learning strategy are the highest among all the learning strategies with values ranging from 12.0 (±1.91) % (MLP) to 14.0 (±3.68) % (Lasso) (Appendix A - Table A.7). Conversely, $\mathrm{MAPE_f}$ obtained in the PerFarm learning strategy are the lowest among the learning strategies with values ranging from 9.0 (±2.0) % (VR) to 10.0 (±2.25) % (kNN) (Figure 4; Appendix A - Table A.8).

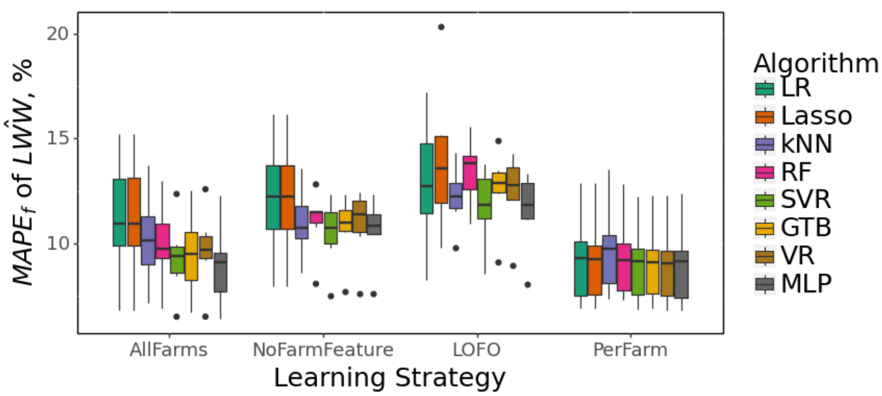

Figure 4: Box plots of mean absolute percentage error per farm ($\mathrm{MAPE_f}$, %) for the predictions of litter weight at weaning (LWW) according to supervised algorithms and learning strategies. For each box plot, the black line represents the median value and whiskers represent 1.5 times the interquartile range (LR, Linear Regression; LASSO, Least Absolute Shrinkage and Selection Operator; kNN, k-nearest-neighbors; RF, Random Forest; SVR, Support Vector for Regression; GTB, Gradient Tree Boosting; VR, Voting Regressor; MLP, Multi Layer Perceptron; AllFarms, all farms learning strategy; NoFarmFeature, no farm feature learning strategy; LOFO, leave-one-farm-out learning strategy; PerFarm, per farm learning strategy)

Learning methods were ranked according to the $\mathrm{MAPE_f}$ metric (Figure 5). For a given learning strategy, the average rank was computed according to the ranks obtained by each algorithm in each farm. The average ranks were 1.33, 2.00, 2.71, and 3.96, for AllFarms, NoFarmFeature, LOFO, and PerFarm learning strategies, respectively, and the critical difference computed according to the Nemenyi test was 0.677 ($P < 0.05$). According to this test, the rankings of PerFarm and AllFarms do not differ significantly, and these learning strategies performed significantly better than the NoFarmFeature and LOFO learning strategies (Figure 5). The rank obtained by the NoFarmFeature was statistically higher than for the LOFO learning strategy, which was at lowest of all.

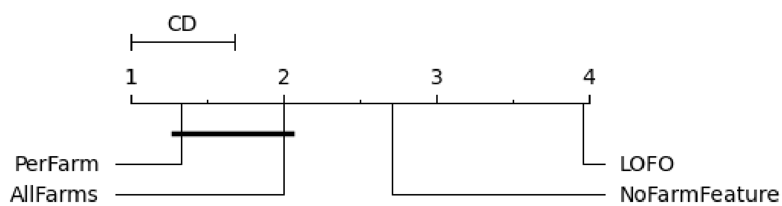

Figure 5: Ranking of learning methods based on their average ranks across datasets and machine learning algorithms, according to the mean absolute percentage error per farm ($\mathrm{MAPE_f}$, %) metric. The black line connects all learning strategies that do not perform statistically differently according to the Nemenyi test.

3.3. Predictions of secondary outcomes

Secondary litter performance criteria and nutrient output in milk were derived from the predictions of LWW obtained with the PerFarm learning strategy, and the VR algorithm, which performed the best. They were first assessed at farm level according to the ME_f metric, which are presented for each farm in table 3. The ME_f obtained for the prediction of LWW and LWG are very similar, and range from -0.59 kg (farm 5) to 0.19 kg (farm 4), compared to average values of 87.04 kg and 71.85 kg for LWW and LWG, respectively. The ME_f obtained for the prediction of LADG range from 30 g/d (farm 5) to 10 g/d (farm 4). The ME_f obtained for the prediction of PWW and PWG range from -0.07 kg (farm 5) to 0.01 kg (farm 3). The ME_f obtained for the prediction of PADG range from -3.69 g/d (farm 5) to 0.16 g/d (farm 4). The ME_f obtained for the prediction of DM_m range from -0.02 kg/d (farm 5) to 0.01 kg/ (farm 4). The ME_f obtained for the prediction of E_m range from -0.61 MJ/d (farm 5) to 0.16 MJ/d (farm 4). The ME_f obtained for the prediction of N_m range from -0.76 g/d (farm 5) to 0.21 g/d (farm 4).

Table 3: Mean value and ME_f of the prediction of LWG, LADG, PWW, PWG, PADG, DM_m, E_m, and N_m derived from the prediction of LWW obtained with the PerFarm learning strategy and the VR algorithm[1]

	LWW kg	LWG kg	LADG kg/d	PWW kg	PWG kg	PADG g/d	DM_m kg/d	E_m MJ/d	N_m g/d
Mean	87.04	71.85	2.58	8.28	6.84	246.79	1.78	49.20	70.76
Farm 1	-0.23	-0.23	-0.01	-0.03	-0.03	-0.93	-0.01	-0.15	-0.19
Farm 2	-0.31	-0.31	-0.01	-0.03	-0.03	-1.02	-0.01	-0.23	-0.29
Farm 3	0.11	0.11	0.00	0.01	0.01	0.15	0.00	0.05	0.06
Farm 4	0.19	0.19	0.01	0.00	0.00	0.16	0.01	0.16	0.21
Farm 5	-0.59	-0.59	-0.03	-0.07	-0.07	-3.69	-0.02	-0.61	-0.76
Farm 6	-0.02	-0.02	0.00	-0.01	-0.01	-0.08	0.00	0.02	0.02
Mean of ME_f	-0.14	-0.14	-0.01	-0.02	-0.02	-0.90	-0.00	-0.13	-0.16
STD of ME_f	0.27	0.27	0.01	0.03	0.03	1.33	0.01	0.25	0.32

[1] ME_f: mean error per farm, LWW: litter weight at weaning, LWG: litter weight gain, LADG: litter average daily gain, PWW: piglet weight at weaning, PWG: piglet weight gain, PADG: piglet average daily gain, DM_m: dry matter in milk, E_m: energy in milk, N_m: nitrogen in milk, VR: Voting Regressor

Predictions of secondary litter performance criteria and nutrient output in milk were then assessed at farm level according to the $\mathrm{MAPE_f}$ metric, and expressed as a percentage of mean value (Table 4). For all criteria, the best predictions were obtained in farm 3 with an average value of 8.0% (±0.7%) across the nine criteria. The quality of predictions was slightly lower for farm 4 and 6 (average MAPE of 8.6% and 10.0%, respectively). The worst prediction was obtained for farm 1 (average MAPE of 13.8%), with farm 2 and 5 falling in the middle (average MAPE of 11.5% and 11.8%, respectively). Mean MAPE across farms was the lowest for LWW and PWW with 9.0% and 9.2%, respectively. For LWG, LADG, PWG, PADG, $\mathrm{DM_m}$, $\mathrm{E_m}$, mean MAPE accross farms was higher (11.3% on average). Mean MAPE across farms fell in the middle for $\mathrm{N_m}$ (10.3%).

Table 4: $MAPE_f$ (%) for LWG, LADG, PWW, PWG, PADG, DM_m, E_m, and N_m derived from the prediction of LWW obtained with the PerFarm learning strategy, and the VR algorithm

	LWW	LWG	LADG	PWW	PWG	PADG	DM_m	E_m	N_m
Farm 1	12.29	14.84	14.77	12.66	15.28	15.21	15.38	15.96	13.85
Farm 2	9.70	11.84	11.83	9.83	11.98	11.97	12.35	12.84	11.05
Farm 3	6.84	8.21	8.20	6.87	8.23	8.23	8.53	8.84	7.70
Farm 4	7.15	8.89	8.89	7.26	8.95	8.94	9.25	9.60	8.34
Farm 5	9.54	12.17	12.29	9.77	12.44	12.57	12.87	13.43	11.42
Farm 6	8.52	10.26	10.24	8.65	10.40	10.37	10.62	10.99	9.65
Mean of $MAPE_f$	9.01	11.03	11.04	9.17	11.21	11.22	11.50	11.94	10.33
STD of $MAPE_f$	1.82	2.22	2.21	1.92	2.36	2.35	2.32	2.42	2.06

[1] $MAPE_f$: mean absolute percentage error per farm, LWW: litter weight at weaning, LWG: litter weight gain, LADG: litter average daily gain, PWW: piglet weight at weaning, PWG: piglet weight gain, PADG: piglet average daily gain, DM_m: dry matter in milk, E_m: energy in milk, N_m: nitrogen in milk, VR: Voting Regressor

4. Discussion

4.1. Factors influencing litter performance criteria and nutrient output in milk

According to the ANOVA, a large proportion of the variance of the criteria related to milk production (*i.e.* LADG, LWW, DM_m, E_m, and N_m) was explained by the effects of farm, parity, litter size at weaning, and month of farrowing. For LADG, 54% of the variance was explained, which is consistent with the results of Ngo et al. (2012b), who obtained a value of 50%. In contrast, the proportion of the variance of litter characteristics at birth (LSB, LWB) or piglet performance (PWB, PWW, PADG) explained by these effects was much lower, with values between 12% and 22%. According to these results, it appears that criteria related to litter performance should therefore be easier to accurately predict than those related to piglet performance. Among the factors affecting the prediction of these parameters, it appears that the number of suckling piglets affecting the number of functional teats explained a greater proportion of the variability of LADG (*i.e.* from 1.3 kg/d for 5 suckling piglets up to more than 3 kg/d for 15 piglets), whereas the effect of parity (about 0.25 kg/d between extreme values) and month of farrowing (about 0.1 kg/d between extreme values) contributed less to explaining its variability.

The LADG increased between sows of first and third parity, and subsequently decreased. Beyer et al. (2007) and Dourmad et al. (2012) also observed that milk production, for which LADG is a proxy, increased between sows of first and second parity, and was the highest for fourth parity sows.

The reduced LADG observed in the summer months, compared to the winter months, may be related to the effects of temperature. The effects of high temperatures on the metabolic activity of the mammary gland are known to be partly due to the reduction in feed intake of sows resulting from heat stress, which consequently reduces the amount of nutrients available for milk synthesis

(Renaudeau et al., 2001). Another finding on the effect of high temperatures was a partial redirection of blood flow to the skin capillaries in order to increase body heat losses (Renaudeau et al., 2003).

In the present data set, LADG strongly changes with litter size, with greater LADG and smaller PADG as litter size increases. LADG linearly increased between 5 and 15 piglets, and tended to decrease after that, whereas PADG showed a curvilinear decrease, with highest values for 5-6 suckling piglets and a large drop for litters with more than 15 piglets at weaning. These results are consistent with Dourmad et al. (2012), who observed that milk production peaked at 12-14 suckling piglets, whereas the amount of milk available per piglet peaked at 7 piglets and decreased linearly above this value. The increase in LADG up to 15 piglets is due, first, to the increase in the number of functional glands, since there is one teat per piglet, and, second, to a stronger stimulation of milk production and a higher suckling intensity (Auldist et al., 1998). On the other hand, the decrease in LADG for litters with more than 15 piglets at weaning might be due to behavioral problems resulting from an imbalance between the number of piglets and the number of functional teats (Orgeur et al., 2004).

The mean and the variability of LADG in the present study showed large differences between the 6 farms. In fact, this farm effect may be linked to other factors that are known to affect litter performance, such as the genetic origin of sows, feeding practices during lactation, farming practices, and environmental conditions (Quesnel et al., 2015). Collecting data on these effects is a rather tedious task, and features representing them were not available in our data set, though they would have been of interest. The farm effect can, however, be of help in examining them through a unique criterion.

4.2. Performance of the different learning strategies

The ME_f of the prediction of LWW showed large differences, depending on the learning strategy. The ME_f obtained with algorithms trained according to the LOFO learning strategy were generally more variable, with values ranging from -10 to +10 kg of litter weight at weaning (*e.g.* between -11% and +11% of the mean LWW), depending on the farm. This clearly indicates that this recursive learning strategy is not accurate, regardless of the algorithm used. The ME_f in the NoFarmFeature learning strategy ranged from -8 to +8 kg of litter weight at weaning. Compared to LOFO, this strategy slightly reduces the occurrence of structural errors between predictions and observations, but is still not accurate enough, regardless of the algorithm used. In contrast, the ME_f obtained with AllFarms and PerFarm were centered on 0, with a very low variability that can be expressed in just hundreds of grams of litter weight at weaning, regardless of the algorithm used and the farm. These strategies, regardless of the data used at learning, showed a great ability to fit and correctly predict LWW with all algorithms. This clearly indicates that the farm feature is of major importance for accurate prediction of LWW. The absence of the farm feature in the LOFO and the NoFarmFeature prediction strategies resulted in inaccurate predictions of average LWW with rather large deviations for some farms.

However, because prediction errors may cancel each other out, the ME_f metric is not sufficient to accurately evaluate the performance of a prediction strategy. In addition to ME, MAPE gives more insights on the extent of prediction errors, both over or under estimations. The $MAPE_f$ for the prediction of LWW also shows large differences, depending on the learning strategy. The Nemenyi test for this metric indicates that the PerFarm and AllFarms learning strategies were ranked similarly over the farms and the algorithms. These two strategies performed better than the NoFarmFeature learning strategy, which in turn performed better than the LOFO learning strategy. This difference in efficiency of learning strategy with respect to MAPE is consistent with what was previously obtained for ME.

Among all the strategies, the lowest $\mathrm{MAPE_f}$ values were obtained in the PerFarm learning strategy, and the results suggest little difference between simple and advanced machine learning algorithms. In the AllFarms learning strategy, low $\mathrm{MAPE_f}$ were also obtained but with great differences depending on the algorithm. With this learning strategy, kNN, SVR, GTB, VR and MLP algorithms better predict the LWW outcome than LR or LASSO algorithms. This indicates that LR and LASSO algorithms are unable to generalize over a dataset that results from multiple Gaussian distributions, even when trained with an indication on the source of each distribution (*i.e.*, the farm name feature). Furthermore, the better performance of complex machine learning algorithms (such as neural-networks and ensemble algorithms) in the AllFarms learning strategy might be due to their ability to model non-linear relationships between features (Warner et al., 2020).

In contrast with the AllFarms learning strategy, using a more complex algorithm in the PerFarm learning strategy only improved performance marginally. It seems that the non-linear relationships between features caught by advanced machine learning algorithms in the AllFarms learning strategy are only induced by the different farms in the training sets. Training algorithms at farm level thus makes it possible to learn more appropriate relationships between target and features, reduces computational time, and leads to more interesting predictive performance than training algorithms over multiple farms. This observation could be compared to clusterwise regression techniques (Gitman et al., 2018), which aim to partition data into clusters before fitting linear regressions per cluster, in order to minimize the overall error. In the case of predicting litter performance during lactation, the underlying model between performance and our set of features looks specific to each farm. These observations are in line with Pietersma et al. (2003) and Warner et al. (2020), who found superior machine learning algorithm performance using small data sets in dairy cow herds, and with Gauthier et al. (2021), who found better performance using time series clustering applied to farm specific data, for further prediction of daily feed intake in lactating sows.

4.3. Prediction of litter performance criteria and nutrient output in milk, for precision livestock farming

The best prediction of LWW, as assessed with the two farm metrics used in this study, show better performance in the PerFarm learning strategy, with small differences according to supervised algorithms. The mean $\mathrm{MAPE_f}$ obtained with the best algorithm is 9.008% (VR algorithm). Considering the small set of features used, this error seems acceptable. The residual part of the variability that was not predictable might reflect the share of variability in the data that was not caught during training with our set of features. This could, for instance, be nutritional features or behavioral features, which are known to affect milk production and consequently litter growth (Orgeur et al., 2004; Etienne et al., 2000). The $\mathrm{MAPE_f}$ also shows farm-specific variations, with extreme values ranging from 6.8 to 12.3% (in the case of the VR algorithm). These differences in residual variability between farms could be due to differences in the period of data collection (in farm 1 data collection occureed between 2000 and 2019, while in farm 4 it only occured in 2017), changes in the genetic origin of sows over the period of data collection, or changes in other factors affecting the milk production. For instance, it would be interesting to consider the delivery of creep feed to piglets, which may vary according to farmer practices, and could help explain the difference between predicted LWW and observed LWW. It would also be interesting to train an algorithm by replacing the LSW feature used in this study with the mean LS across the lactating period. This could give better insight into milk production, by taking into account the events that may change LS and LWW, such as deaths and cross fostering of piglets.

Predicting secondary outcomes is necessary for implementing precision feeding systems for lactating sows in practice (Gauthier et al., 2019), or for any other application relying on the prediction of a proxy for milk production. When using the best learning strategy (PerFarm) and the best supervised learning algorithm (VR), ME values do not differ from zero for LWW as well as for the secondary outcomes related to milk production. The $MAPE_f$ of secondary outcomes show only a very small reduction compared to the $MAPE_f$ of LWW (11% for LADG, E_m and N_m on average, compared to 9% for LWW), indicating that the precision of this technique is nearly as good as for LWW. Thus, the prediction of secondary outcomes offers almost the same predictive performance as the prediction of LWW, with only a very limited decrease in predictive performance compared to LWW. With this method, training models for only LWW can be used to reliably predict multiple outcomes derived from this criteria.

4.4. Use of the algorithm in practice

This study highlights key elements for embedding an algorithm able to predict litter performance and nutrient output in milk in decision support systems for the precision feeding of lactating sows. This algorithm provides individual estimations of LWW and nutrient outputs in milk in lactating sows, thus making it possible to evaluate daily nutrient requirements for each individual lactating sow (Dourmad et al., 2008; NRC, 2012; Gauthier et al., 2019). From the perspective of precision feeding, the present results indicate that a limited number of features can be sufficient, but the algorithm should be trained at farm level. Aggregating data from different farms does not significantly improve performance. It is suggested that training the algorithms at farm level helps take into account factors influencing milk production and litter performance related to environmental conditions, animal behavior, or farming practices that are not easy to measure.

As for computational cost and the prediction performance achieved by the different algorithms, it is suggested that the LASSO and LR algorithms are the best suited for obtaining fast and reliable predictions of litter performance and nutrient output in milk. The loss of interpretability and the computational cost that advanced machine learning algorithms such as GTB, VR, and MLP entail, are not offset by improved prediction in that situation. Moreover, from a practical point of view, these simple algorithms might be trained on-farm, on a standard PC, with no dedicated infrastructure needed, and the prediction equations could be easily embedded in an on-farm decision support system. This is even more important as predictive performance may decrease when farming practices or sow genetics evolve within the farm, thus requiring training the algorithm with new data. LWW and secondary outcomes can thus be predicted with few predictors and good reliability, while requiring computational resources available at the farm level.

5. Conclusion

Based on previous studies that provided good descriptions of the factors influencing milk production, litter performance criteria, and nutrient output in milk in lactating sows, the present study showed that these factors could be used and easily collected on-farm in order to build simple yet reliable predictive models of litter weight at weaning, at the farm level. Secondary outcomes used by precision feeding decision support systems can be accurately obtained from the predictions of litter weight at weaning, and used to provide accurate estimations of individual nutrient output in milk for nutrition decision support systems for lactating sows.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge the following partners that provided the data used in this study: Chambres d'agriculture de Bretagne (Rennes, France), Laval University (QC, Quebec), CDPQ (QC, Quebec), IFIP (Le Rheu, France), INRAE GenESI (Le Magneraud, France), and INRAE UE3P (Saint Gilles, France). The authors also gratefully acknowledge the GenOuest bioinformatics core facility (https://www.genouest.org) for providing the computing infrastructure, and Corentin Raphalen for his help at early stage of this study. This study formed part of a Ph.D. thesis in the #DigitAg project (ANR-16-CONV-0004), supported by the French National Research Agency in the "Investments for the Future" program; and the European Union's Horizon 2020 Research and Innovation program (grant agreement no. 633531).

References

Auldist, D.E., Morrish, L., Eason, P., King, R.H., 1998. The influence of litter size on milk production of sows. Animal Science 67, 333–337. URL: https://www.cambridge.org/core/product/identifier/S1357729800010109/type/journal_article, doi:10.1017/S1357729800010109.

Bascol, K., Emonet, R., Fromont, E., Debusschere, R., 2017. Improving Chairlift Security with Deep Learning, in: Adams, N., Tucker, A., Weston, D. (Eds.), Advances in Intelligent Data Analysis XVI. Springer International Publishing, Cham. volume 10584, pp. 1–13. URL: http://link.springer.com/10.1007/978-3-319-68765-0_1, doi:10.1007/978-3-319-68765-0_1.

Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. J. Mach. Learn. Res. 13, 281–305.

Beyer, M., Jentsch, W., Kuhla, S., Wittenburg, H., Kreienbring, F., Scholze, H., Rudolph, P.E., Metges, C.C., 2007. Effects of dietary energy intake during gestation and lactation on milk yield and composition of first, second and fourth parity sows. Archives of Animal Nutrition 61, 452–468. URL: http://www.tandfonline.com/doi/full/10.1080/17450390701563433, doi:10.1080/17450390701563433.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., Vanderplas, J., Joly, A., Holt, B., Varoquaux, G., 2013. API design for machine learning software: experiences from the scikit-learn project URL: http://arxiv.org/abs/1309.0238.

Demšar, J., 2006. Statistical Comparisons of Classifiers over Multiple Data Sets. J. Mach. Learn. Res. 7, 1–30.

Dourmad, J., Étienne, M., Valancogne, A., Dubois, S., van Milgen, J., Noblet, J., 2008. InraPorc: A model and decision support tool for the nutrition of sows. Animal Feed Science and Technology 143, 372–386. URL: http://linkinghub.elsevier.com/retrieve/pii/S0377840107001770https://linkinghub.elsevier.com/retrieve/pii/S0377840107001770, doi:10.1016/j.anifeedsci.2007.05.019.

Dourmad, J.Y., Quiniou, N., Heugebaert, S., Paboeuf, F., Ngo, T.T., 2012. Effect of parity and number of suckling piglets on milk production of sows, in: Book of Abstracts of the 63rd Annual Meeting of the European Association for Animal Production. Wageningen Academic publishers, Bratislava, Slovakia. volume 18 of *EAAP Book of abstracts*, p. 44. URL: http://www.wageningenacademic.com/9789086867615, doi:10.3920/978-90-8686-761-5.

Etienne, M., Legault, C., Dourmad, J.Y., Noblet, J., 2000. Production laitière de la truie: Estimation, composition, facteurs de variation et évolution. Journées de la Recherche Porcine 32, 253–264. URL: http://www.journees-recherche-porcine.com/texte/2000/00txtAlim/A0013.pdf.

Gaillard, C., Brossard, L., Dourmad, J.Y., 2020. Improvement of feed and nutrient efficiency in pig production through precision feeding. Animal Feed Science and Technology 268, 114611. URL: https://linkinghub.elsevier.com/retrieve/pii/S0377840120305150, doi:10.1016/j.anifeedsci.2020.114611.

Gauthier, R., Largouët, C., Gaillard, C., Cloutier, L., Guay, F., Dourmad, J.Y., 2019. Dynamic modeling of nutrient use and individual requirements of lactating sows. Journal of Animal Science 97, 2822–2836. URL: https://academic.oup.com/jas/advance-article/doi/10.1093/jas/skz167/5494821https://academic.oup.com/jas/article/97/7/2822/5494821, doi:10.1093/jas/skz167.

Gauthier, R., Largouët, C., Rozé, L., Dourmad, J.Y., 2021. Online forecasting of daily feed intake in lactating sows supported by offline time-series clustering, for precision livestock farming. Computers and Electronics in

Agriculture 188. URL: https://linkinghub.elsevier.com/retrieve/pii/S016816992100346X, doi:10.1016/j.compag.2021.106329.

Géron, A., 2019. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media.

Gitman, I., Chen, J., Lei, E., Dubrawski, A., 2018. Novel Prediction Techniques Based on Clusterwise Linear Regression URL: http://arxiv.org/abs/1804.10742.

Hansen, A.V., Strathe, A.B., Kebreab, E., France, J., Theil, P.K., 2012. Predicting milk yield and composition in lactating sows: A Bayesian approach. Journal of Animal Science 90, 2285–2298. URL: https://academic.oup.com/jas/article/90/7/2285/4702030, doi:10.2527/jas.2011-4788.

Ngo, T.T., Quiniou, N., Heugebaert, S., Paboeuf, F., Dourmad, J.Y., 2012a. Influence du rang de portée et du nombre de porcelets allaités sur la production laitière des truies, in: 44. Journées de la recherche porcine, IFIP-Institut du Porc. p. np. URL: http://prodinra.inra.fr/ft?id=%7B9B4E5FC1-810C-4324-AC83-D5F6E44A0AE8%7D.

Ngo, T.T., Quiniou, N., Heugebaert, S., Paboeuf, F., Dourmad, J.Y., 2012b. Influence du rang de portée et du nombre de porcelets allaités sur la production laitière des truies. 44. Journées de la recherche porcine , npURL: http://prodinra.inra.fr/ft?id=%7B9B4E5FC1-810C-4324-AC83-D5F6E44A0AE8%7D.

Noblet, J., Dourmad, J.Y., Etienne, M., 1990. Energy utilization in pregnant and lactating sows: modeling of energy requirements. Journal of animal science 68, 562–572. URL: https://academic.oup.com/jas/article/68/2/562-572/4631828, doi:10.2527/1990.682562x.

Noblet, J., Etienne, M., 1989. Estimation of sow milk nutrient output. Journal of animal science 67, 3352–3359. URL: https://www.animalsciencepublications.org/publications/jas/abstracts/67/12/JAN0670123352https://academic.oup.com/jas/article/67/12/3352-3359/4697124, doi:10.2527/jas1989.67123352x.

NRC, 2012. Nutrient Requirements of Swine. 11th rev. ed., Natl. Acad. Press, Washington, DC.

Orgeur, P., Le Dividich, J., Saez, E., Salaun, C., Le Roux, T., 2004. La taille de la portée influe sur le comportement des porcelets à la mamelle et sur leur croissance. Journees De La Recherche Porcine En France 36, 457.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Pietersma, D., Lacroix, R., Lefebvre, D., Wade, K.M., 2003. Performance analysis for machine-learning experiments using small data sets. Computers and Electronics in Agriculture 38, 1–17. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168169902001047, doi:10.1016/S0168-1699(02)00104-7.

Pomar, C., van Milgen, J., Remus, A., 2019. Precision livestock feeding, principle and practice. Poultry and Pig Nutrition. Challenges of the 21st Century, 397–418.

Quesnel, H., Farmer, C., Theil, P.K., 2015. Colostrum and milk production, in: Farmer, C. (Ed.), The gestating and lactating sow. Wageningen Academic

Publishers, Wageningen, The Netherlands. chapter 8, pp. 173–192. doi:https://www.wageningenacademic.com/doi/10.3920/978-90-8686-803-2_8.

Renaudeau, D., Noblet, J., Dourmad, J.Y., 2003. Effect of ambient temperature on mammary gland metabolism in lactating sows. Journal of animal science 81, 217–231. URL: https://dl.sciencesocieties.org/publications/jas/abstracts/81/1/217.

Renaudeau, D., Quiniou, N., Noblet, J., 2001. Effects of exposure to high ambient temperature and dietary protein level on performance of multiparous lactating sows. Journal of Animal Science 79, 1240. URL: https://academic.oup.com/jas/article/79/5/1240-1249/4682894, doi:10.2527/2001.7951240x.

Seabold, S., Perktold, J., 2010. statsmodels: Econometric and statistical modeling with python, in: 9th Python in Science Conference.

Silalahi, P., Tribout, T., Prunier, A., Billon, Y., Gogué, J., Bidanel, J.P., 2016. Estimation of the effects of selection on French Large White reproductive performance using frozen semen. Journal of Animal Science 94, 3655–3662. URL: https://academic.oup.com/jas/article/94/9/3655/4701682, doi:10.2527/jas.2016-0540.

Vranken, E., Berckmans, D., 2017. Precision livestock farming for pigs. Animal Frontiers 7, 32–37. URL: https://academic.oup.com/af/article/7/1/32/4638771, doi:10.2527/af.2017.0106.

Warner, D., Vasseur, E., Lefebvre, D.M., Lacroix, R., 2020. A machine learning based decision aid for lameness in dairy herds using farm-based records. Computers and Electronics in Agriculture 169, 105193. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168169919319969, doi:10.1016/j.compag.2019.105193.

Wathes, C., Kristensen, H., Aerts, J.M., Berckmans, D., 2008. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? Computers and Electronics in Agriculture 64, 2–10. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168169908001476, doi:10.1016/j.compag.2008.05.005.

Appendix A. Mean Absolute Percentage Errors

In this appendix, the mean absolute percentage errors at farm level, obtained for a given learning strategy and all supervised learning algorithms, are presented. MAPE per farm is provided, along with its mean value, and its standard deviation across farms.

Table A.5: Mean absolute percentage error per farm (%) for the prediction of litter weight at weaning, according to the AllFarms learning strategy, for each algorithm[1]. Best score per line is in boldface.

	LR	Lasso	kNN	RF	SVR	GTB	VR	MLP
Farm 1	13.501	13.532	13.740	12.965	12.361	12.533	12.587	**12.312**
Farm 2	9.987	9.985	10.694	10.125	9.664	9.740	**9.651**	9.695
Farm 3	6.821	6.825	7.180	6.928	6.510	6.721	6.513	**6.419**
Farm 4	15.223	15.223	8.781	9.248	8.459	7.911	9.771	**7.274**
Farm 5	11.919	11.913	11.516	11.194	9.927	10.805	10.537	**9.071**
Farm 6	9.883	9.883	9.606	9.387	**9.129**	9.228	9.181	9.160
Mean of $MAPE_f$	11.222	11.227	10.253	9.974	9.342	9.490	9.707	**8.989**
STD of $MAPE_f$	2.719	2.723	2.081	1.853	**1.751**	1.883	1.801	1.876

[1] LR: Linear Regression, LASSO: Least Absolute Shrinkage and Selection Operator, kNN: k-nearest neighbors, RF: Random Forest, SVR: Support Vector for Regression, GTB: Gradient Tree Boosting, VR: Voting Regressor, MLP: Multi-Layer Perceptron

Table A.6: Mean absolute percentage error per farm (%) for the prediction of litter weight at weaning, according to the NoFarmFeature learning strategy, for each algorithm[1]. Best score per line is in boldface.

	LR	Lasso	kNN	RF	SVR	GTB	VR	MLP
Farm 1	13.127	13.128	13.589	12.817	12.317	12.329	12.452	**12.315**
Farm 2	10.504	10.500	11.107	10.818	10.438	10.555	**10.351**	10.457
Farm 3	7.975	7.971	8.613	8.078	**7.500**	7.735	7.610	7.612
Farm 4	13.932	13.928	**10.391**	11.532	11.618	11.702	12.143	11.427
Farm 5	16.153	16.144	10.210	11.467	**9.824**	10.727	11.644	10.442
Farm 6	11.356	11.359	11.996	11.488	**11.080**	11.276	11.139	11.202
Mean of $MAPE_f$	12.174	12.172	10.984	11.034	**10.463**	10.721	10.890	10.576
STD of $MAPE_f$	2.611	2.610	1.549	**1.450**	1.546	1.461	1.617	1.469

[1] LR: Linear Regression, LASSO: Least Absolute Shrinkage and Selection Operator, kNN: k-nearest neighbors, RF: Random Forest, SVR: Support Vector for Regression, GTB: Gradient Tree Boosting, VR: Voting Regressor, MLP: Multi-Layer Perceptron

Table A.7: Mean absolute percentage error per farm (%) for the prediction of litter weight at weaning, according to the Leave-One-Farm-Out learning strategy, for each algorithm[1]. Best score per line is in boldface.

	LR	Lasso	kNN	RF	SVR	GTB	VR	MLP
Farm 1	13.675	14.872	14.347	15.573	13.761	14.886	14.256	**13.312**
Farm 2	11.283	12.282	12.149	14.282	11.151	13.487	12.107	**11.134**
Farm 3	8.277	9.800	9.822	10.963	8.546	9.094	8.955	**8.073**
Farm 4	15.129	15.181	**11.555**	13.785	13.324	13.111	13.687	13.074
Farm 5	17.192	20.344	12.316	12.216	**11.346**	12.389	13.433	12.222
Farm 6	11.834	11.822	13.077	13.834	12.323	12.656	12.099	**11.426**
Mean of $\mathrm{MAPE_f}$	12.898	14.050	12.211	13.442	11.742	12.604	12.423	**11.540**
STD of $\mathrm{MAPE_f}$	2.861	3.360	**1.382**	1.481	1.714	1.761	1.742	1.740

[1] LR: Linear Regression, LASSO: Least Absolute Shrinkage and Selection Operator, kNN: k-nearest neighbors, RF: Random Forest, SVR: Support Vector for Regression, GTB: Gradient Tree Boosting, VR: Voting Regressor, MLP: Multi-Layer Perceptron

Table A.8: Mean absolute percentage error per farm (%) for the prediction of litter weight at weaning, according to the PerFarm learning strategy, for each algorithm[1]. Best score per line is in boldface.

	LR	Lasso	kNN	RF	SVR	GTB	VR	MLP
Farm 1	12.873	12.872	13.542	12.829	**12.250**	12.274	12.291	12.378
Farm 2	9.954	9.936	10.513	10.229	9.685	9.805	9.698	**9.666**
Farm 3	6.930	6.915	7.651	7.367	6.855	6.934	6.841	**6.822**
Farm 4	7.142	7.172	7.342	7.292	7.240	7.251	7.150	**7.002**
Farm 5	10.170	9.844	10.015	**9.333**	9.793	9.405	9.543	9.617
Farm 6	8.685	8.704	9.531	9.046	8.583	8.779	**8.522**	8.678
Mean of $\mathrm{MAPE_f}$	9.292	9.241	9.766	9.349	9.067	9.075	**9.008**	9.027
STD of $\mathrm{MAPE_f}$	2.025	2.000	2.053	1.877	1.803	**1.773**	1.822	1.874

[1] LR: Linear Regression, LASSO: Least Absolute Shrinkage and Selection Operator, kNN: k-nearest neighbors, RF: Random Forest, SVR: Support Vector for Regression, GTB: Gradient Tree Boosting, VR: Voting Regressor, MLP: Multi-Layer Perceptron