Online forecasting of daily feed intake in lactating sows supported by offline time-series clustering, for precision livestock farming

Raphaël Gauthier, Christine Largouët, Laurence Rozé, Jean-Yves Dourmad

Computers and Electronics in Agriculture 188 (2021) 106329

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Online forecasting of daily feed intake in lactating sows supported by offline time-series clustering, for precision livestock farming

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

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

ARTICLE INFO

Keywords:
Feed intake
Lactating sow
k-Shape clustering
Time-series forecasting
Data Mining
Precision Livestock Farming

ABSTRACT

According to precision livestock farming principles, it is essential to apply feed intake forecasting processes to real time precision feeding strategies in order to improve the overall efficiency of the livestock feeding chain. Considering the lack of a mechanistic model that predicts daily feed intake in lactating sows, a novel approach combining an online forecasting procedure with an offline learning procedure is proposed. A database of 39,090 lactations, from 6 different farms and containing the first 20 daily feed intake records after farrowing, was used (1) to identify consistent sets of clusters and trajectory curves offline, and (2) to test 3 predictive functions of daily feed intake online. The homogeneity of the clusters resulting from the offline learning procedure was assessed according to Silhouette and Calinski-Harabasz scores. The predictive quality of forecasting functions was assessed with the Mean Error (ME), and the Root Mean Square Error (RMSE). Time-series clustering with *k*-Shape makes it possible to extract consistent trajectory curves that are scale-, shift- and translate-invariant. The best number of clusters obtained either in a global approach or at farm scale was two. The trajectory curve of the first cluster is characterized by a mostly continuous increase of feed intake over the course of lactation, and the second cluster by a plateau in feed intake starting from about the 10th day of lactation. These identified trajectory curves are consistent with the very few studies available in the literature. When computed with the best forecasting function and farm specific trajectory curves, the ME of feed intake over lactation was −0.08 kg/d, and the corresponding RMSE was 1.06 kg/d. Though variability in feed intake among sows and over the lactation period is high, online forecasting of feed intake can be improved by the use of feed intake trajectory curves. These trajectory curves may be computed on a regular basis with data obtained directly on the farm or on farms with similar practices. The online forecasting procedure requires few computing resources, and could easily be embedded in smart feeder control systems as a practical application in precision feeding systems for lactating sows.

1. Introduction

Feeding is an essential component of livestock production systems with respect to animal health and welfare, farm sustainability, and competitiveness. The overall efficiency of the livestock feeding chain is largely dependent on the match between nutrient supply and animal requirements, in order to limit nutrient wastage while achieving production objectives. In practice, all pigs at a given physiological stage are generally fed with the same standard diet corresponding to the requirements of an average animal representing the population. However, according to precision livestock farming principles (Vranken and Berckmans, 2017), addressing the diversity among animals could be an effective lever in building more efficient feeding systems (Pomar et al., 2019; Gaillard et al., 2020). With the help of smart feeders, new sensors, and information technology, modern precision feeding systems for growing pigs have demonstrated their ability to meet individual requirements more efficiently (Cloutier et al., 2015).

In lactating sows, high milk production and low voluntary feed intake generally lead to nutrient deficiency (Noblet et al., 1990), especially in primiparous sows. To limit nutrient deficiency, one common practice consists in supplying *ad libitum* access to a feed with a high nutrient content, with the risk of increasing feeding cost and nutrient

* Corresponding author.
E-mail address: raphael.gauthier@gmx.com (R. Gauthier).

https://doi.org/10.1016/j.compag.2021.106329
Received 27 October 2020; Received in revised form 10 June 2021; Accepted 12 July 2021
Available online 5 August 2021

excretion, and, consequently, reducing overall sustainability. Precision feeding has not yet been evaluated for lactating sows but seems to be a promising strategy with respect to the large variability of nutrient requirements among sows (NRC, 2012; Gauthier et al., 2019). To operate in real-time, precision feeding systems need to accurately predict the feed intake of lactating sows on a daily and individual basis in order to adjust the optimal mix between two diets, one with a high nutrient content and the second with a low nutrient content, as already described by Pomar et al. (2019) for growing pigs. As reviewed by O'Grady et al. (1985) and Eissen et al. (2000) sow voluntary feed intake during lactation is affected by many factors such as sow's parity, body weight and backfat thickness at farrowing, and litter size. Feed intake is also very sensitive to ambient temperature with a negative effect of hot conditions that has been quantified by Ribeiro et al. (2018) in a meta-analysis, and modeled on a daily basis by Staicu et al. (2020) and Cabezón et al. (2016). However, these studies generally, only, report the average lactation feed consumption or average daily feed intake curves (Schinckel et al., 2010). Thus, the factors affecting the variability in the pattern of daily feed consumption over lactation have been investigated to a lesser extent, unless in the study from Koketsu et al. (1996) who identified different patterns of intake over manually collected feed intake data, and the study from Cabezón et al. (2017) who investigated lactating sows feed intake patterns using a statistical approach based on polynomial prediction functions.

With the current development of sensors and computing resources, huge amounts of data are automatically and continuously collected in the form of discrete or continuous measurements, images, videos, and sounds. Valuable knowledge can be extracted from this data with adapted machine learning techniques. Among all data types, time-series have become pervasive in recent decades, with active research work and applications in many different fields such as stock market analysis, weather forecasts, and power consumption monitoring. Thus, the recent development of innovative feeders providing access to daily and individual feed consumption of sows, as well as the development of specific computational methods for storing and dealing with time-series (Aghabozorgi et al., 2015) are providing new opportunities to develop more accurate predictions of feed intake.

In this study, we present a forecasting procedure for time-series supported by unsupervised learning of consistent clusters, specifically designed to make one-day-ahead forecasts of sow feed intake during lactation. Using data from different farms, our approach first uses time-series clustering to automatically identify consistent sets of feed intake trajectory curves (TCs) during lactation. Then, three functions are tested to make one-day-ahead forecasts of individual sow feed intake, with two of them supported by an assigned TC. The objective of this study is thus to describe and assess the quality of this approach to predicting individual feed intake in lactating sows.

2. Research background

With the increasing amount of time-series data in various domains, temporal data mining has recently attracted a great deal of attention for different purposes such as classification, visualization, segmentation, prediction, and trend analysis, in addition to pattern discovery. Time-series clustering is one of the most fundamental task that is usually applied prior to any other analysis method. This section briefly reviews specific knowledge about time-series clustering in order to extract temporal prototypes. These prototypes will further be used to support time-series forecasting.

2.1. Time-series definitions

We begin by introducing the necessary definitions.

Definition 1. A time-series T consists of a sequence of numerical vectors in successive order and equally spaced out over time: $T = t_1, t_2, \cdots, t_n$, with t_i being a V-dimensional real-valued vector and n the length of the time-series T.

The time-series T is *univariate* when $V = 1$, meaning that only one variable varies over time, otherwise T is *multivariate*. In this paper, we deal with univariate time-series, a sequence of numerical values over time.

Definition 2. A dataset D is a set of time-series such that: $D = \{T_1, T_2, \cdots, T_m\}$, where m is the number of time-series in the dataset.

2.2. Time-series clustering

In machine learning, clustering belongs to the class of *unsupervised learning* problems whose objective is to determine how the data is organized without any labeled examples. The objective of clustering is to partition the dataset into homogeneous groups of data, called *clusters*, where data points in the same cluster are the more similar to each other and dissimilar to data in other clusters. Clustering can be applied to time-series and is defined as follows (Aghabozorgi et al., 2015):

Definition 3. Given a time-series dataset $D = \{T_1, T_2, \cdots, T_m\}$, time-series clustering consists in partitioning D into $C = \{C_1, C_2, .., C_k\}$, a set of k clusters, with $D = \bigcup_{i=1}^{k} C_i, C_i \cap C_j = \varnothing$ for $i \neq j$. Homogeneous time-series are grouped together based on a certain similarity measure that maximizes inter-cluster distance and minimizes intra-cluster variance.

The main challenges of the clustering process are to define *similarity* and to find the value of k that leads to a consistent set of clusters.

2.2.1. Time-series clustering algorithms

Numerous approaches have been proposed to deal with time-series objects characterized by large data sizes and potentially high dimensionality. For whole time-series clustering (as opposed to subsequence clustering, for instance), clustering algorithms are generally classified into three groups, namely shape-based, feature-based, and model-based, depending on whether or not clustering is applied directly to raw data (Warren Liao, 2005; Aghabozorgi et al., 2015).

Feature-based and model-based algorithms are not directly applicable to raw data and require time-series conversion. Feature-based algorithms work on vectors of features extracted from raw time-series, such as mean, variance, autocorrelation, etc. (Bandara et al., 2020). This leads to dimensionality reduction, thus making it possible to cluster datasets that cannot fit into memory and time-series of unequal lengths. These algorithms are therefore generally less computationally expensive (Aghabozorgi et al., 2015). Model-based algorithms first model each time-series, for instance with an Auto-Regressive Moving Average (ARMA) model, for instance, and the clustering is carried out on the parameters of the obtained models (Warren Liao, 2005; Aghabozorgi et al., 2015). Shape-based clustering algorithms operate directly on raw data, and time-series that share a common progression across time are grouped together. These types of algorithms are good at capturing redundant patterns over time since they rely on a measure of similarity/dissimilarity specially designed for time-series. Depending on the number, length, and dimensionality of the time-series being compared, and the complexity of that measure, shape-based algorithms can lead to high computational cost.

Paparrizos and Gravano (2016) proposed a scalable and efficient shape-based clustering algorithm, called k-Shape, that uses a normalized version of the cross-correlation measure as its distance measure. This algorithm can effectively detect similarities in time-series presenting invariances such as scaling, shift, and translation (Fig. 1). k-Shape is based on a two-step iterative procedure, which shares similarities with the procedure of the well-known k-Means algorithm. First, each time-series is assigned to the cluster for which the similarity between time-series and the cluster's centroid is greatest. Then, the centroid is

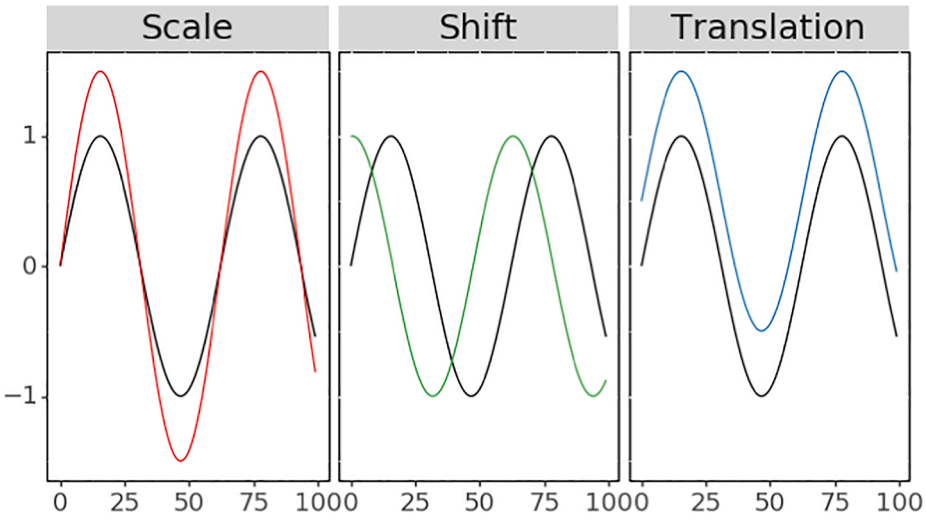


Fig. 1. Illustration of scaling, shift, and translation invariances applied to a sinusoidal function (black curve). Despite different distortions, it might be interesting to consider the similarity in shape of the red, green, and blue lines with the black line.

computed again for each cluster to reflect changes in cluster membership. In the case of the *k*-Shape algorithm, the centroid is an artificial sequence. The algorithm is initialized by randomly assigning a time-series to one of the clusters and is stopped when no more changes occur in cluster assignment or when a maximum number of iterations is reached. Scaling and translation invariances are handled by *z*-normalizing each time-series before applying the *k*-Shape algorithm, so its mean is 0 and its standard deviation is 1. Shift invariances are handled by the Shape-Based Distance, which is presented in Section 2.2.2.

2.2.2. Distance measures for shape-based clustering algorithms

Measuring the similarity/dissimilarity between time-series is a major step in clustering algorithms and is usually carried out with a distance measure. Due to the temporal aspect of this data and the different complexities emerging from the various domains, many distance measures have been proposed in the literature (Wang et al., 2013).

The most common distance measure in time-series is the Euclidean distance, given by the following formula:

$$dist(T_1, T_2) = \sqrt{\sum_{i=1}^{n} (T_{1i} - T_{2i})^2}$$

where T_1 and T_2 are two time-series of equal length n. Because the Euclidean distance computes the square differences of observations sharing the same time index, this measure is quite fast. However, it is not well suited to comparing time-series of unequal lengths, or presenting shift and translation invariances (Fig. 1).

These drawbacks can be compensated by *elastic measures* that compare the local alignment of time-series independently of the time index. For example, Dynamic Time Warping (DTW, Sakoe and Chiba, 1978) can accurately identify the similarity of time-series presenting temporal drifts or varying in lengths by comparing one value in T_1 with T_2 in three different ways, namely one-to-one, one-to-many or one-to-none. DTW is thus more accurate than Euclidean distance and is considered to be the best distance measure for many time-series mining tasks (Ding et al., 2008; Bagnall et al., 2017), but it is also much slower and more computationally expensive.

When the objective of time-series clustering is to identify common trajectory curves or behaviors in the object of interest, shaped-based distances are better suited to comparing trajectory shapes. To circumvent the main drawback of classic shape-based distance metrics like *Fréchet* and *Hausdorff* that are computationally expensive, a new SBD measure (Shape-Based Distance) has recently been proposed by Paparrizos and Gravano (2016). SBD relies on cross-correlation, a statistical measure that makes it possible to compare the shapes of two time-series T_1 and T_2 of unequal lengths n and m by reducing their noise (Aghabozorgi et al., 2015). The SBD distance measure also handles shift invariances.

Definition 4. SBD distance is defined as follows:

$$SBD(T_1, T_2) = 1 - \max_{w}(NCC_w(T_1, T_2))$$

where $NCC_w(T_1, T_2) = (ncc_1, \cdots, ncc_w), w \in \{1, 2, \cdots, n + m - 1\}$, is the normalized cross-correlation sequence. The normalized cross-correlation sequence $NCC_w(T_1, T_2) = (ncc_1, \cdots, ncc_w)$ is computed for all w positions obtained by keeping one time-series static and sliding the other over it (Paparrizos and Gravano, 2016). SBD is computed at the position w that maximizes the similarity between T_1 and T_2. The SBD distance measure then varies between 0 and 2, where 0 indicates that T_1 and T_2 are perfectly similar. The time requirement for computing the normalized cross-correlation sequence for all w values is high, particularly for long time-series, but this drawback is handled by using Fast Fourier Transform (Paparrizos and Gravano, 2016).

2.2.3. Cluster prototypes

Clustering makes it possible to automatically identify relevant groups of time-series without any *a priori* knowledge on cluster definition. An additional task in time-series clustering consists in computing a "prototype" for each cluster. Prototypes computed during the clustering process are used directly by some clustering algorithms (*i.e. k*-Medoids, *k*-Shape) to refine cluster membership. Prototypes computed at the end of the clustering process offer a single representative time-series for each cluster that can be used for further applications such as time-series forecasting. Prototypes are either a medoid or a centroid sequence. A medoid sequence is an actual time-series from the cluster, while a centroid sequence is an artificial time-series computed from the time-series of the cluster (Aghabozorgi et al., 2015).

3. Materials and Methods

3.1. General approach

The general approach of this study aims to define groups of sows having the same feed intake trajectory curve (offline learning through time-series clustering) to support the forecasting of the individual daily feed intake of lactating sows (online time-series forecasting). This

approach is based on the combination of the individual lactating sow online data acquired since farrowing with herd historical data collected during previous farrowing batches (Fig. 2).

A feed intake time-series $F_j = f_{1,j}, f_{2,j}, \cdots, f_{n,j}$ is a sequence of numerical values $f_{d,j}$ that represent the feed intake value in kilograms at day d, $d \in [1, n]$, where n is the duration of the lactation period for a sow j.

The offline learning procedure (Fig. 2) performed on herd historical data consists in a clustering algorithm that splits past recorded F_j time-series into homogeneous clusters (see Section 2). A prototype is then extracted in order to summarize the feed intake trajectory curve (TC) followed by each cluster. This offline learning requires the availability of sufficient data to be able to extract consistent prototypes.

The online forecasting (Fig. 2) consists in predicting the value $\hat{f}_{d+1,j}$ of the time-series $F_j, d \in [1, n] \cap \mathbb{N}$. Forecasting uses both an assigned TC to sow j and $f_{1,j}, f_{2,j}, \cdots, f_{d,j}$ sow live data recorded since farrowing. The forecast $\hat{f}_{d+1,j}$ may then be transmitted to any application that relies on daily feed intake forecasts at an individual level (*e.g.* precision feeding). The amount of feed really consumed, called $f_{d+1,j}$, is finally recorded from the feeder by the end of the day in order to be used on the following days. Data preprocessing and offline and online methods are described in the following subsections.

3.2. Data preprocessing

Data collection. Data was collected using an automated feeder (Gestal®, JYGA Technologies Inc., Quebec, Canada) that recorded the feed intake and the feeding behavior of lactating sows on a daily basis. With this system, sows were delivered up to 8 meals over the course of the day. Feed was given in successive portions of a limited size, which were distributed by the feeder when the sows pressed a button. This allowed the sows to be fed according to their demand, while limiting feed wastage effectively. Each sow was assigned to a predefined feeding scale depending on parity, with a daily target consumption and permission given by the farmer to exceed this target by up to 30%. Each sow could therefore ask for less than the target, but never more than the maximum.

Data cleaning. Daily feed intake was recorded between April 2013 and June 2019 in 6 commercial farms where the farrowing crates were equipped with Gestal® system. The original database was composed of 78,863 lactations of variable lengths. Lactations shorter than 12 days or longer than 32 days were withdrawn, as they may represent abnormal events (*e.g.* early death of sow after farrowing, adoption of a second litter). It was assumed that occasional electronic anomalies occurred when encountering missing or negative daily feed intake records, and daily feed intake higher than 6 kg on first day after farrowing, greater than 15 kg on the second day or greater than 20 kg from day 3 on. It was also assumed that several electronic anomalies occurred while recording feed intake within a lactation when the cumulative feed intake during lactation was over 250 kg. When one of these situations arose, the whole lactation was removed from the dataset. In addition, the last day of lactation showed a distinctive bimodal distribution, with some sows showing a huge drop in feed intake. Because this drop was likely related to specific feed allowance practices on weaning day (*i.e.*, feeding only half of the ration), this day was excluded from offline learning and online forecasting. The combination of these cleaning steps led to a clean database of 64,951 lactations.

Data selection and data splitting. Most lactations last between 17 and 23 days as a result of biological variability in the duration of the gestational period and batch management at weaning (Martel, 2008). Shorter lactations may be due to the occurrence of lactation problems, whereas longer lactations may be related to specific practices such as keeping sows longer to nurse fostered piglets. In this study, for the selection of time-series, we used 20-day lactations as a compromise

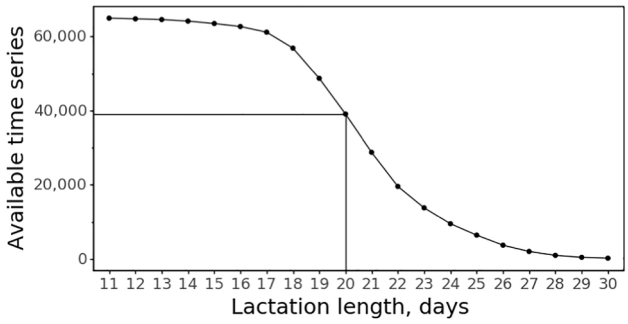


Fig. 3. Number of feed intake time-series available according to the length of the lactation period.

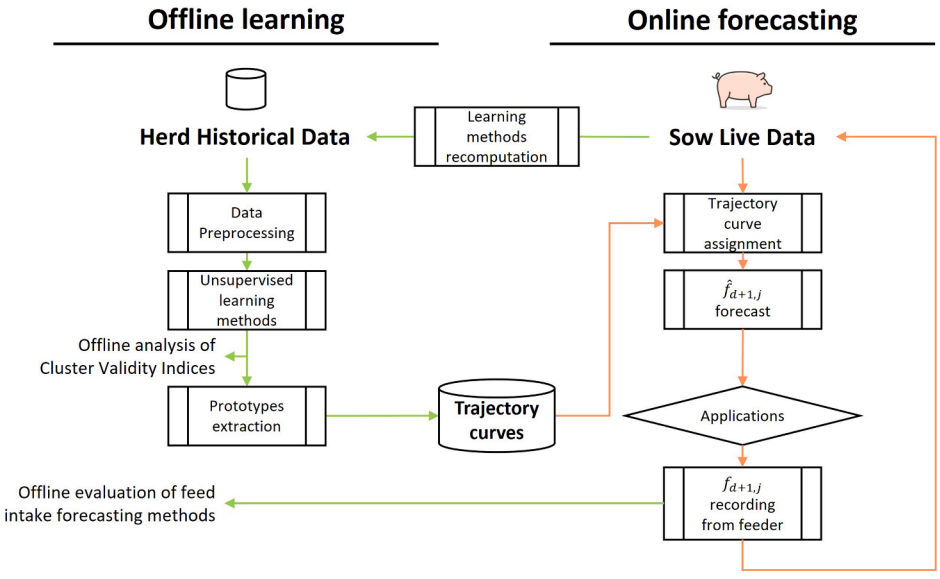


Fig. 2. Methodological approach for the daily prediction of sow feed intake during lactation, from individual sow live data and herd historical data. $\hat{f}_{d+1,j}$ represents the forecasted value of feed intake for sow j on day $d+1$. $f_{d+1,j}$ represents the amount of feed really consumed at the end of day $d+1$ for the sow j. Blue lines represent offline steps and red lines represent online forecasting tasks for feed intake time-series.

between decreasing the number of available F_j time-series, and increasing the lactation length to cover a wider range of practices (Fig. 3). This resulted in the selection of 39,090 20-day F_j time-series. This dataset, called D, was then split at random into training and test sets according to an 80:20 ratio, and with respect to this ratio in each farm. The training set of D was used for offline learning of feed intake trajectory curves (TCs), both per farm in the farm specific (FS) approach, and over all farms in the global (G) approach. The test set of D was used for the validation of the online forecasting simulation.

Data description. The number of time-series available per farm, mean feed intake, and the standard deviation for each of the 6 farms in dataset D are presented in Table 1. Mean feed intake differed among farms. It was greatest in farm 2 and lowest in farm 1 and averaged 6.05 kg (±1.29). Over the period of 20 days, feed intake showed a gradual increase (Fig. 4). During the five first days of lactation, mean feed intake increased rapidly from 2 kg to 5 kg, and subsequently continued to increase more slowly, in line with the sows' appetite, reaching a plateau in the third week of lactation. The variability in feed intake was large and increased over lactation as the result of the wide biological variability in appetite between individual sows and between successive days (Eissen et al., 2000). For the same reason there was also an increasing number of outliers calculated as being higher or lower than the interquartile range multiplied by 1.5.

3.3. Unsupervised learning and prototype extraction

As one of the more efficient time-series clustering algorithm, the *k*-Shape clustering algorithm was used to identify different sets C of feeding trajectory curves (TCs) in the training set of D (see 2.2.1) (Paparrizos and Gravano, 2016).

F_j time-series have a lot of variability (Fig. 4), and may present scaling, translation, and shift invariances. *k*-Shape handles scaling and translation invariances by *z*-normalizing time-series, so the mean and standard deviation values of each time-series in the training set were first set to 0 and 1, respectively. This step was necessary in order to identify similarity in the feeding behavior of sows despite possible differences in their feed intake level. *k*-Shape handles shift invariance thanks to Shape-Based Distance (SBD). This clustering algorithm is thus able to identify the similarity of different time-series with a common progression of f_j occurring at different time indices. Prototype extraction relies on the Shape Extraction algorithm (Paparrizos and Gravano, 2016). *k*-Shape was also chosen because of its ability to deal with numerous time-series of equal length, and its domain-independent nature (Paparrizos and Gravano, 2016).

k-Shape algorithm takes only one input parameter, which is k, the desired number of clusters. In this study, k represents the number of clusters in which time-series shared a common feed intake TC. Since this value cannot be known *a priori*, we made k vary between 2 to 8, and analyzed the homogeneity of the resulting clusters with the Silhouette (Rousseeuw, 1987) and Calinski-Harabasz (Calinski and Harabasz, 1974) scores. The Silhouette score is an internal Cluster Validity Index (internal CVI) that evaluates the homogeneity of time-series within the cluster and the heterogeneity between clusters. The Silhouette score varies between −1 and + 1, where the value + 1 indicates that clusters

Table 1
Number of feed intake time-series per farm, and means and standard deviation of average 20-days lactation feed intake.

Farm	Number of time series	Mean feed-intake (kg/d)
1	7 872	5.14 ± 0.865
2	3 467	6.98 ± 1.422
3	9 111	6.50 ± 1.119
4	651	6.22 ± 1.132
5	10 692	6.08 ± 1.260
6	7 297	5.99 ± 1.279
All	39 090	6.05 ± 1.294

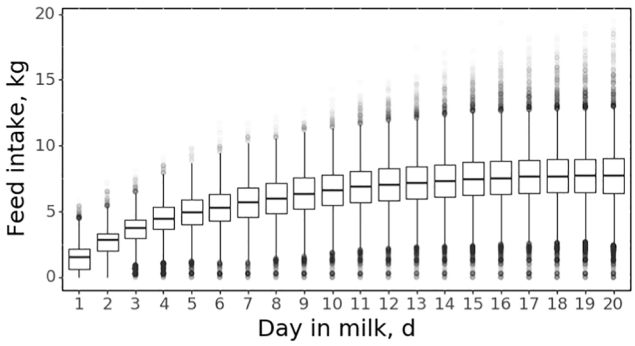


Fig. 4. Daily feed intake boxplots according to lactation stage of sows (data obtained from 39090 time-series in six commercial farms).

are well separated by the clustering algorithm and that time-series within a cluster are very similar, and the value −1 indicates the opposite situation. The Calinski-Harabasz score is a second CVI, which is defined as the ratio between the within-cluster dispersion and the between-cluster dispersion. Higher score indicates that clusters are well separated.

Offline processes were carried out both per farm and globally over the training set. The first approach computed global (G) prototypes given the whole training set of D, while the second approach worked by farm and computed farm-specific (FS) prototypes. In total, the FS approach produced 7 sets of clusters C_k per farm, $k \in [2, 8]$, and the G approach produced 7 sets of $C_k, k \in [2,8]$. All sets of clusters were tested with Silhouette and Calinski-Harabasz scores. Each prototype was turned into a trajectory curve TC expressed in kilograms per day by applying back the original mean and standard deviation values gathered during z-normalization.

3.4. One-day-ahead forecasting of feed intake

In this section, we present the general principles of the online forecasting procedure, the TC assignment task, the 3 forecasting functions used, and the evaluation of predicted feed intakes (Table 2).

General principles. The one-day-ahead forecast of feed intake is denoted $\widehat{f}_{d+1,j}$, on lactating day d+1, for a specific sow j belonging to the test set of D. Online forecasting starts at day 2 and is computed each day for each sow, from both an assigned TC and the previous f_d of the sow recorded since farrowing (Fig. 2). Two forecasting functions were based on the previous f_d of the sow and benefited from offline learning of feed intake TCs. A third baseline forecasting function was performed

Table 2
Summary of the one-day-ahead forecasting methods, with different offline and online parameters, and the resulting method short name used in the text.

Offline learning method	Online forecasting function	k[a]	Method short name
Global clustering[b] with *k*-Shape	1	[2,8]	*G.1f.k*
Farm-specific clustering[c] with *k*-Shape	1	[2,8]	*FS.1f.k*
Global clustering with *k*-Shape	2	[2,8]	*G.2f.k*
Farm-specific clustering with *k*-Shape	2	[2,8]	*FS.2f.k*
No offline learning	3		*Persistence*

[a] k, the clustering parameter, is equivalent to the number of trajectory curves available for assignment during online forecasting
[b] Global (G) clustering computes k trajectory curves, given the whole training database.
[c] Farm-specific (FS) clustering computes k trajectory curves, given the training database of each farm.

exclusively online. All TCs derived from G and FS prototypes $\forall k \in [2,8]$ were tested in the subsequent online forecasting procedure of F_j time-series. All forecasting methods are summarized in Table 2.

Trajectory curve assignment. The time-series F_j was assigned to the $TC \in C_k, k \in [2,8]$, which shared the most similar progression of F_j since farrowing. This similarity is evaluated with SBD in z-normalized conditions. All throughout the lactation period, the F_j could be assigned to different TCs.

Forecasting functions. Two forecasting functions were tested after TC assignment. The first (1f) computes a single one-day-ahead forecast $\widehat{f}_{d+1,j}$, which corresponds to the last observed $f_{d,j}$ value of F_j increased by the variation of the assigned prototype TC_j such that:

$$\widehat{f}_{d+1,j} = f_{d,j} + (TC_{d+1,j} - TC_{d,j}) \quad (1)$$

The second (2f) computes two forecasts, a one-day-ahead forecast from d, and a two-days-ahead forecast from d - 1 and returns the average value at $d+1$. This function is used to mitigate the possible variability of $f_{d,j}$ from one day to another:

$$\widehat{f}_{d+1,j} = \frac{(f_{d,j} + (TC_{d+1,j} - TC_{d,j})) + (f_{d-1,j} + (TC_{d+1,j} - TC_{d-1,j})}{2} \quad (2)$$

A third baseline forecasting function was performed exclusively online, to evaluate and compare the benefits of TCs in feed intake forecasting. This method, called "Persistence" (Table 2), is a naive forecasting baseline where the forecast at $d+1$ corresponds to the last observed $f_{d,j}$ value. Therefore, this forecasting function does not benefit from any herd historical data:

$$\widehat{f}_{d+1,j} = f_{d,j} \quad (3)$$

Error measures and quality evaluation. To evaluate the precision of the forecasting methods and to identify the method with the best predictive quality, errors between the forecast $\widehat{f}_{d+1,j}$ and the ground truth $f_{d+1,j}$ were analyzed. Mean Error per lactating sow j (ME_j) was computed such that:

$$ME_j = \frac{1}{20}\sum_{d=1}^{20} \widehat{f}_{d,j} - f_{d,j} \quad (4)$$

A positive or a negative ME_j indicates that the predictive method tends to overestimate or underestimate F_j over the lactation period of sow j.

Root Mean Square Errors were computed to both evaluate the effects of daily variability ($RMSE_d$) and sow variability ($RMSE_j$) on the quality of the prediction. $RMSE_d$ evaluates the progression of the predictive quality of each method according to lactation stage. RMSE_d, $d \in [1,20]$ were computed each day over the 7,818 time-series in the test set of D such that:

$$RMSE_d = \sqrt{\frac{1}{7818}\sum_{j=1}^{7818} (\widehat{f}_{d,j} - f_{d,j})^2} \quad (5)$$

A $RMSE_d$ value close to 0 indicates very good forecasts on day d. $RMSE_j$ evaluates the precision of predictive methods according to individual sows. $RMSE_j$, $j \in [1,7818]$, were computed such that:

$$RMSE_j = \sqrt{\frac{1}{20}\sum_{d=1}^{20} (\widehat{f}_{d+1,j} - f_{d+1,j})^2} \quad (6)$$

A $RMSE_j$ value close to 0 indicates very good forecasts over the whole lactation period of sow j. $RMSE_j$ makes it possible to evaluate which method has good predictive quality for a specific herd.

4. Results

4.1. Offline cluster identification, prototype, and trajectory curve extraction

In the FS approach, offline learning identified 7 sets of clusters per farm, and extracted for each cluster the corresponding prototype and trajectory curve for each cluster. In the G approach, offline learning identified 7 sets of clusters over all 6 farms, and extracted for each cluster the corresponding prototype and trajectory curve. Computed Silhouette and Calinski-Harabasz scores for all sets of clusters are presented in Table 3. The Silhouette score was always maximum for $k = 2$ in each of the 6 farms (FS approach) and also across all the time-series in the training set of D (G approach). For $k = 2$, the Silhouette scores varied from 0.16 up to 0.22. For $k \in [3,5]$, the Silhouette scores decreases consistently, and for $k \in [6,8]$, the Silhouette scores were lower than 0.10. In the FS approach, the Calinksi-Harabasz scores were maximal for $k = 2$ in farms 1, 2, 4, 5, and 6. For farm 3, this score was maximum for $k = 3$. In the G approach, this score was maximal for $k = 2$. Based on these observations, $k = 2$ was chosen as the best parameter to split the training set of D, both in the FS and G approaches, into a consistent set of two clusters.

The Fig. 5 represents each cluster, obtained for $k = 2$, with its z-normalized prototype. Within each cluster, the corresponding prototype is a smooth and artificial time-series that averages all the time-series.

4.2. Evaluation of online forecasting methods

Table 4 presents the ME and RMSE errors per day for methods FS.1f.k, FS.2f.k, G.1f.k, G.2f.k, $k \in [2,8]$, and Persistence.

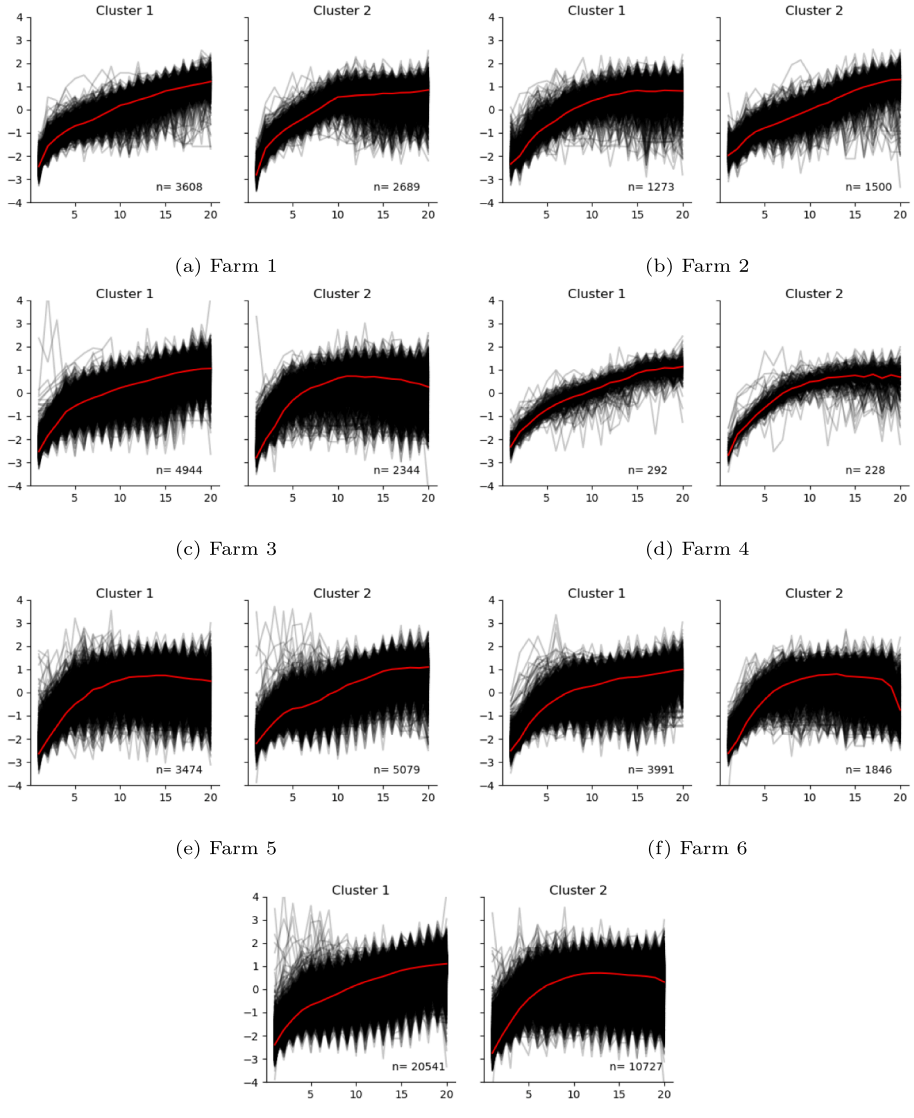
For all methods supported by offline learning of TCs, the smallest ME_j values were achieved for $k = 2$ (Table 4). In the FS.1f.k method, ME_j was equal to -0.08 kg/d for $k = 2$ and decreased down to -0.15 kg/d for $k = 7$. In the FS.2f.k method, ME_j was equal to -0.08 kg/d for $k = 2$ and decreased down to -0.14 kg/d for $k = 8$. In the G.1f.k method, ME_j was equal to -0.04 kg/d for $k = 2$ and decreased down to 0.09 kg/d for $k = 8$. In the G.2f.k method, ME_j was equal to -0.04 kg/d for $k = 2$ and decreased down to -0.12 kg/d for $k = 7$. In the Persistence method, ME_j was equal to -0.31 kg/d. The distribution of ME_j values among sows is presented in Fig. 6 for the four cluster-based methods, with $k = 2$, and for the Persistence method. The smallest ME_j was obtained for G.1f.2 method (Fig. 6). With this method, 75% of the sows had an ME value between -0.10 and $+0.05$ kg/d.

For all methods supported by offline learning of TCs, the smallest $RMSE_j$ were obtained for $k = 2$ (Table 4). In the FS.1f.k method, the $RMSE_j$ was smallest for $k = 2$ (1.11 kg/d) and increased up to 1.21 kg/d for $k = 7$. In the FS.2f.k method, the $RMSE_j$ was smallest for $k = 2$ (1.06 kg/d) and increased up to 1.21 kg/d for $k = 7$. In the G.1f.k method, the smallest $RMSE_j$ was obtained for $k = 2$ with 1.13 kg/d and was maximal for $k = 8$ with 1.19 kg/d. In G.2f.k, the $RMSE_j$ was smallest for $k = 2$ with 1.07 kg/d and increased up to 1.16 kg/d for $k = 5$. In comparison, the $RMSE_j$ obtained with the Persistence method was 1.21 kg/d. The distribution of $RMSE_j$ values among sows is presented in Fig. 7 for the four cluster-based methods, with $k = 2$, and for the Persistence method. The smallest $RMSE_j$ was obtained with FS.2f method associated with $k = 2$ (Fig. 7). With this method, 75% of the sows had a mean $RMSE_j$ value between 0.75 and 1.3 kg/d.

Fig. 8 presents the $RMSE_d$ errors expressed as a percentage of the mean values for the four cluster-based methods, with $k = 2$, and the Persistence method. For all methods, $RMSE_d$ decreased over the first 5 days of lactation and then plateaued at a low level. On day one, $RMSE_{d=1}$ was smallest for FS.1f.2 and FS.2f.2 and represented 45.6% of the average true f_d. For G.1f.2, G.2f.2 and Persistence, the $RMSE_{d=1}$ was higher and reached 69.3% of the average true f_1. From day 1 to 3, $RMSE_d$ quickly decreased and reached about 20% for all of cluster-based methods, while it remained greater at 32.5% with the Persistence

Table 3
Evaluation of the quality of clustering according to k, the number of clusters and consequently the number of trajectory curves produced during offline learning. Best values for Silhouette (maximum) and Calinski-Harabasz (maximum) scores in boldface.

		k						
	Farm	2	3	4	5	6	7	8
Silhouette score	1	**0.18**	0.10	0.10	0.13	0.04	0.03	0.02
	2	**0.16**	0.11	0.07	0.05	0.03	0.03	0.01
	3	**0.22**	0.10	0.06	0.07	0.04	0.04	0.02
	4	**0.17**	0.09	0.09	0.04	0.06	0.01	−0.00
	5	**0.19**	0.10	0.05	0.04	0.04	0.05	0.04
	6	**0.22**	0.19	0.12	0.08	0.08	0.08	0.08
	All	**0.22**	0.10	0.10	0.06	0.06	0.04	0.01
Calinski-Harabasz score	1	**403**	258	347	308	279	247	221
	2	**102**	95	85	88	73	61	58
	3	358	**600**	492	468	397	351	349
	4	**21**	16	17	12	11	11	12
	5	**412**	349	303	284	239	250	219
	6	**488**	445	361	316	297	260	269
	All	**1539**	1305	1213	1053	943	875	804



(a) Farm 1 (b) Farm 2

(c) Farm 3 (d) Farm 4

(e) Farm 5 (f) Farm 6

(g) All Farms

Fig. 5. Comparison of clusters and z-normalized prototypes for each of the 6 farms and all farms together, identified for $k = 2$.

Table 4
Evaluation of mean error (ME_j) and root mean square error ($RMSE_j$) per sow according to the combinations of learning methods (FS: Farm Specific; G: Global), forecasting functions (1f,2f, Persistence), and the number of clusters $k, k \in [2, 8]$

	k	-	2	3	4	5	6	7	8
ME_j (kg/d)	FS.1f	-	−0.08	−0.10	−0.10	−0.12	−0.11	−0.15	−0.12
	FS.2f	-	−0.08	−0.11	−0.09	−0.13	−0.12	−0.14	−0.14
	G.1f	-	−0.04	−0.05	−0.08	−0.08	−0.08	−0.08	−0.09
	G.2f	-	−0.04	−0.06	−0.09	−0.05	−0.09	−0.12	−0.08
	Persistence	−0.31	-	-	-	-	-	-	-
$RMSE_j$ (kg/d)	FS.1f	-	1.11	1.12	1.13	1.15	1.15	1.21	1.20
	FS.2f	-	1.06	1.09	1.11	1.12	1.13	1.21	1.20
	G.1f	-	1.13	1.14	1.16	1.18	1.18	1.18	1.19
	G.2f	-	1.07	1.09	1.12	1.16	1.13	1.13	1.15
	Persistence	1.21	-	-	-	-	-	-	-

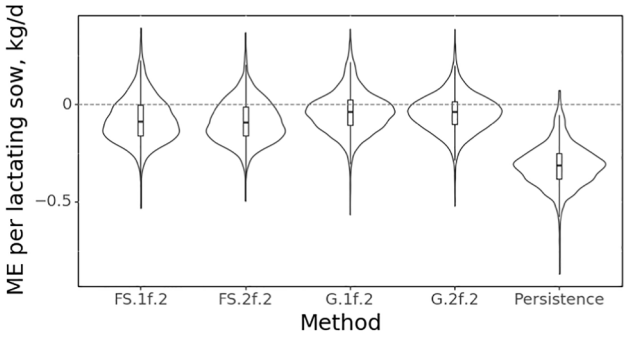


Fig. 6. Mean Error (ME) per lactating sow according to the combination of learning methods (FS: Farm Specific; G: Global) and forecasting functions (1f,2f) with two clusters ($k = 2$), and without learning with the Persistence forecasting function.

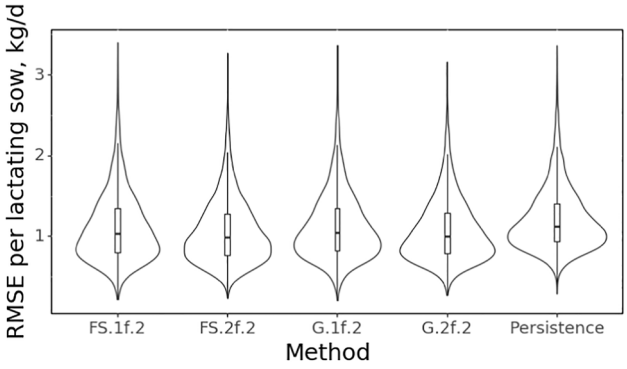


Fig. 7. Root Mean Square Error (RMSE) per lactating sow according to the combination of learning methods (FS: Farm Specific; G: Global) and forecasting functions (1f,2f) with two clusters ($k = 2$), and without learning with the Persistence forecasting function.

method. From day 5 to 20, the $RMSE_d$ remained almost constant. The mean $RMSE_d$, $d \in [5, 20]$, for Persistence, FS.1f.2, G.1f.2, FS.2f.2, and G.2f.2 methods were 18.8%, 18.5%, 18.6%, 17.2%, and 17.2%, respectively.

The effect of lactation stage on ME_d is presented in Fig. 9. For FS.1f.2, FS.2f.2, G.1f.2, and G.2f.2, ME_d was generally negative in the first days of the lactation period (day 1 to day 3), then slightly positive for a few days, and negative again until the end of the lactation. In comparison, the ME_d errors obtained with the Persistence method were almost always negative with huge errors on the first days of the lactation period. ME_d ranged between -0.30 and 0.20 kg/d for the FS.1f.2, FS.2f.2, G.1f.2, and G.2f.2 methods. It ranged between -1.20 kg/d and 0.10 kg/d for the Persistence method.

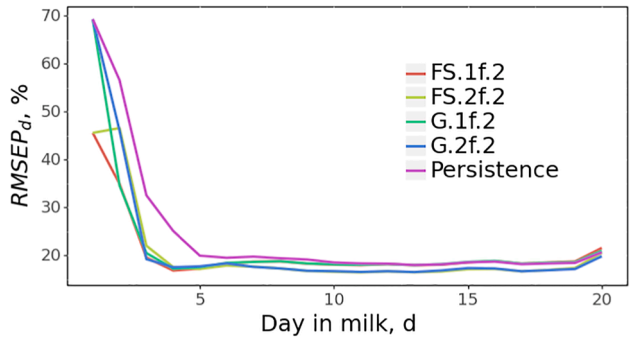


Fig. 8. Effect of lactation stage on RMSEP per day ($RMSE_d$ expressed as a % of the measured value of daily feed intake) according to the combination of learning methods (FS: Farm Specific; G: Global) and forecasting functions (1f,2f) with two clusters ($k = 2$), and without learning with the Persistence forecasting function.

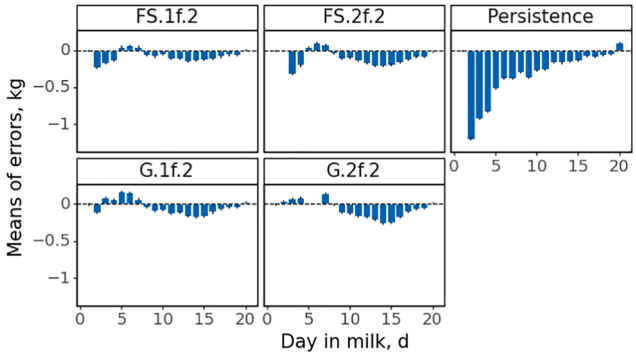


Fig. 9. Effect of lactation stage on Mean Error according to the combination of learning methods (FS: Farm Specific; G: Global) and forecasting functions (1f,2f) with two clusters ($k = 2$), and without learning with the Persistence forecasting function.

5. Discussion

5.1. Offline learning

Clustering was used to split the training set into a consistent set of k clusters. The value of $k = 2$ was found to maximize the Silhouette score in each of the 6 farms and also at the global scale. The Silhouette score is strictly positive (0.20 on average), indicating that time-series tend to be closer to their own cluster than other clusters, but this score is closer to 0 than 1, which indicates that clusters may overlap at some periods (Rousseeuw, 1987). Indeed, in the different clusters, the progression of feed intake from farrowing to day 5 is quite similar, indicating that most sows increase their feed intake in the same way. This might be related to

the progressive increase in their nutrient requirements due to increasing milk production (NRC, 2012; Gauthier et al., 2019) and the progressive adaptation of their digestive tracts (Theil, 2015). This can also be related to the feeding practices, since maximum feed allowance is generally limited over the first 4 to 7 days of lactation in order to avoid digestive disturbances, which are frequent during this period (Göransson, 1989). Although we did not have access to the information about the strategy used in each farm during this period, it seems that daily feeding supply or intake also increased in most studied farms, from about 2.5 kg on day one up to 4 to 5 kg on day five (Fig. 10). Using $k > 2$ did not result in more consistent sets of clusters, as the Silhouette score decreases almost linearly. The Calinski-Harabasz score also confirms that the optimal number of clusters was two. Clustering with *k*-Shape leads to very good handling of shift invariances, as shown by Fig. 5 under normalized conditions.

Though residual variability in *z*-normalized time-series is high, *k*-Shape clustering suggests that time-series could be classified into a limited number of prototypes, independently of feed intake level and variability of individual sows. One prototype found in each of the 6 farms and in the global approach describes a rapid increase during the first 7 days, followed by a slower and almost linear increase. It represented 54 % to 68 % of the time-series in the training set, depending on the farm. The second prototype describes a curvilinear increase of feed intake during the first 7 days, followed by a plateau, starting from around day 10. This second prototype represented between 32 % and 46 % of the time-series in the training set, depending on the farm.

Although the general shapes of the *z*-normalized feed intake patterns (Fig. 5) were rather similar among farms, some specificities could be identified. The variability was much lower in farm 4, probably in relation to the lower number of data available. A feeding pattern with a plateau was observed in all farms but it started earlier in some of them (farm 3 and 4) and was sometime followed by a curvilinear decrease like in farms 3 and 6. These differences could be related the feeding practices in each farm, especially the shape of the target feeding curve programmed in the feeder. It could also be related to some animal specificities, such as the genetic origin, as shown by Schinckel et al. (2010) who compared three breeds of sows and observed that in one breed feed intake plateaued whereas in the two others it continued to increase until weaning.

Fig. 10 represents, for each individual farm and all farms together, the trajectory curves obtained using their two specific prototypes ($k = 2$), after applying back the original conditions of means and standard deviations of the farms to the extracted prototypes. Smooth and farm specific feed intake trajectory curves were thus obtained within each farm, and for all farms together. There are very few studies available in the literature on variability in individual sow feed intake patterns during lactation in commercial swine herds, mainly because feed intake is very rarely recorded, except with the use of smart feeders in recent years (Piñeiro et al., 2019). In the study of Koketsu et al. (1996), daily sow feed intake was measured manually on a large number of sows, with data available for about 25,000 lactations from 30 commercial farms. Average lactation length (19 days) was similar to the present study and average feed intake (5.2 kg/d) was 15% lower than in the present study. They identified six patterns of daily feed intake according to the amount of feed consumed, how quickly feed intake increased, and whether a transient drop in feed intake occured during lactation. Three of these patterns presented similar trends to the two identified in the present study, with either a rapid or a gradual increase in feed intake over time. However, two of the patterns identified in Koketsu et al. (1996)' study, those with a rapid increase and a major or a minor transient drop in feed intake, were not identified in the present dataset, even when considering more than two clusters (results not presented). Dourmad et al. (1991), who observed a drop in feed intake at about five days of lactation in lean sows with high appetite fed *ad libitum* from the day of farrowing, suggested that this was related to the occurrence of gastrointestinal disorders resulting from uncontrolled excessive feed intake at the beginning of lactation. It can be argued that in the present study the use of smart feeders, which make it posible to limit the risk of overconsumption, could have decreased the frequency of such a feeding pattern. According to Koketsu et al. (1994), the "rapid" or "gradual" feeding patterns, which correspond to the two identified in the present study should be encouraged in order to optimize reproduction and lactation performance, and reduce the risk of reproductive failure after weaning. It is thus possible that with another database from farms with different feeding practices (*e.g.* with *ad libitum* feeding since farrowing) other feeding patterns, with a transient drop in feed intake, could be identified. Similar average feed intake patterns, as these identified in the present study, with a rapid or a more gradual increase in feed intake over lactation were also found by Schinckel et al. (2010) who compared three different breeds of sows using a generalized Michaelis–Menten functions to adjust the feed intake curves. As in the present study the daily feed intake of sows increased rapidly from d 1 to 4 of lactation and thereafter increased at a decreasing rate to reach a plateau at about d 18 to 23 of lactation, the level of the plateau being affected by the breed of sows and the season. With longer lactations (27 days) Cabezón et al. (2016), using a Mixed model polynomial functions to adjust the feeding curves, identified some feeding patterns showing a decrease in feed intake over the last week of lactation, as we also observed in some farms from our study. According to the results available in the literature and the results we obtained, it appears that there is a large diversity of feeding patterns in lactating sows. This highlights the importance of regularly carrying out the offline machine learning procedure with data obtained directly from the farm or from other farms with practices similar to those used in the farm applying the forecasting procedure.

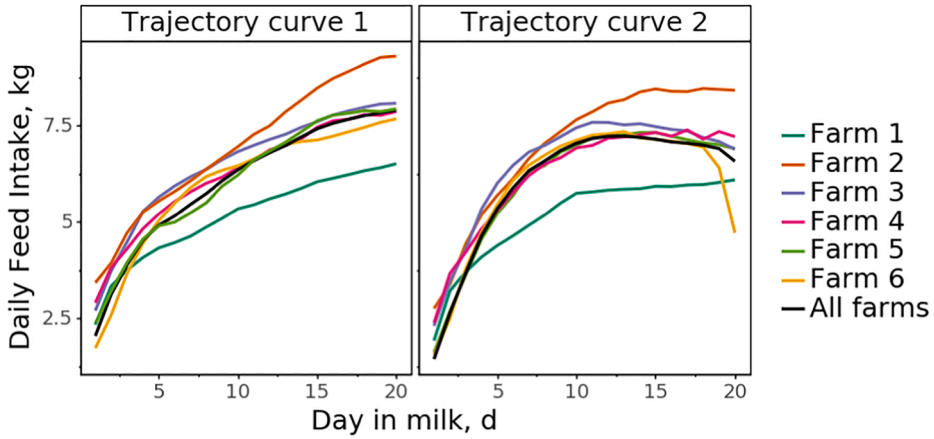


Fig. 10. Comparison of trajectory curves for each of the 6 farms and all farms together, identified for $k = 2$.

5.2. Online forecasting

Online forecasting started with the assignment of a trajectory curve (TC) for each sow and each day based on the Shape-Based Distance between TC and $f_{d,j}$ values. On average, for $k = 2$, the assignment of a given sow to a prototype changed 2.2 times during her lactation. Those changes mainly occurred at the very beginning or at the very end of the lactation period. This might be due to difficulty of comparing smooth extracted TCs with raw individual sow time-series, when the shape changes.

For each of the cluster-based forecasting methods, predictive quality was the best for low values of k. Increasing the number of prototypes decreased predictive quality, probably by assigning a less consistent TC on a given day, thus increasing the number of prototypes changes over lactation. This result may seem counterintuitive since it might be expected that increasing the number or prototypes would improve the prediction.

The approach with the Persistence function and no offline learning was used as a baseline method for comparison with methods supported by offline learning. Fig. 9 clearly shows that this function does not efficiently predict daily feed intake, especially over the first 11 days of lactation. Over this period, the 1f and 2f forecasting functions performed much better than the Persistence function with smaller ME values almost centered on 0 kg and smaller RMSE. After 11 days of lactation, when feed intake is more constant, the ME values were quite comparable for the different forecasting functions. However, because ME errors, which may be positive or negative, may cancel out each other between days, this single criterion is not sufficient to evaluate the accuracy of predictions. *RMSE* thus provides another understanding of the predictive quality. According to both *ME* and *RMSE* criteria, the farm specific method with two prototypes and a forecasting function based on the previous two days' feed intakes (FS.2 f2) appears to be the most suitable (ME_j = -0.08 kg/d, $RMSE_j$ = 1.06 kg/d), although the same method based on all farm data together (G.2 f2) is very close (ME_j = -0.04 kg/d, $RMSE_j$ = 1.07 kg/d).

The smaller RMSEP obtained for the 2f functions compared to the 1f functions $\widehat{f}_{d+1,j}$ is probably due to an improved forecast of the change in feed intake by taking the means between two forecasts. This could explain why the difference in $RMSE_j$ between the 1f and 2f forecasting methods, expressed as a percentage of the mean value, are greater in the beginning of lactation.

5.3. Use of the full approach in practice

As stated in Section 3.1, $\widehat{f}_{d+1,j}$ forecasts may be used by any applications that rely on individual prediction of daily feed intake during lactation at the individual level. In the precision feeding approach, prediction of feed intake is required to determine the optimal nutrient content of the diet that will be prepared by the smart feeder and fed to each individual sow (Gauthier et al., 2019).

Due to k-Shape efficiency and the small length of feed intake time-series, time-series clustering is quite fast and mainly depends on the number of time-series involved in the procedure. It may require less than 3 s to compute k clusters with their corresponding prototypes over 1000 feed intake time-series with a single-core processor; this represents about one year of data in a herd with 400 sows. Conversely, the computation of Silhouette and Calinski-Harabasz scores is far more expensive in time and resources.

Similarly, the online forecasting methods were also very fast and required less than one second to run, for one sow over one day, and require only a few computing resources. The online prediction could thus be easily embedded in smart feeder control systems.

To start the forecasting system on a new farm, where feed intake data are not yet available, global prototypes obtained from the present study might be used. When sufficient feed intake data becomes available on the farm (*i.e.* about 1,000 time-series, as in farm 4), the offline learning procedure could be run in order to produce farm-specific prototypes. This offline learning requires very few parameters and, according to the present results, exploring $k < 4$ seems to offer reasonable guidance. However, careful preprocessing of the data is required, and the approach needs to be combined with automatic detection of electronic anomalies in input data (*e.g.* feeder disconnection, bad data transmission) and correction. This would secure both the offline learning and the online forecasting of feed intake.

6. Conclusion

Forecasting of animal feed intake time-series is a challenging task, with many applications in practice for precision feeding using smart feeders. This approach is the first contribution that applies big data methods to lactating sow feed intake time-series, based on both historical and live data. Clustering with k-Shape makes it possible to extract consistent prototypes and trajectory curves that are scale-, shift-, and translate-invariant. With the data used in this study, k-Shape clustering suggests that time-series could be classified into a limited number of prototypes, despite the fact that feed intake is highly variable during the lactation period. Due to shape-based clustering, our approach is easily interpretable by farmers who are already used to handle the concept of feeding curves when programming their sow feeding systems. Finally, this decision support system might be easily embedded on-farm, for the precision feeding of lactating sows, with few requirements in computing resources, and is able to learn by itself from farm specific data in a machine learning way.

Study involving animals

The data used in this paper were obtained from commercial farms using commercial feeding devices.

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 JYGA Technologies (Quebec, Canada) and the farmers who provided the data used in this study, and Thomas Dahmen 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

Aghabozorgi, S., Seyed Shirkhorshidi, A., Ying Wah, T., 2015. Time-series clustering – A decade review. Information Systems 53, 16–38. https://doi.org/10.1016/j.is.2015.04.007 https://linkinghub.elsevier.com/retrieve/pii/S0306437915000733, arXiv:1107.3326.

Bagnall, A., Lines, J., Bostrom, A., Large, J., Keogh, E., 2017. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Min. Knowl. Disc. https://doi.org/10.1007/s10618-016-0483-9 arXiv:1602.01711.

Bandara, K., Bergmeir, C., Smyl, S., 2020. Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. Expert Syst. Appl. 140, 112896. https://doi.org/10.1016/j.eswa.2019.112896 https://linkinghub.elsevier.com/retrieve/pii/S0957417419306128.

Cabezón, F.A., Schinckel, A.P., Leon, Y.L., Craig, B.A., 2017. Analysis of lactation feed intakes for sows with extended lactation lengths. Translational. Animal Science 1, 1–25. https://doi.org/10.2527/tas2017-0016 https://academic.oup.com/tas/article/1/1/1/5479522.

Cabezón, F., Schinckel, A., Richert, B., Stewart, K., Gandarillas, M., Peralta, W., 2016. Analysis of lactation feed intakes for sows including data on environmental

temperatures and humidity. The Professional Animal. Scientist 32, 333–345. https://doi.org/10.15232/pas.2015-01495 http://linkinghub.elsevier.com/retrieve/pii/S1080744616300171.

Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Communications in Statistics - Theory and Methods 3, 1–27. https://doi.org/10.1080/03610927408827101 http://www.tandfonline.com/doi/abs/10.1080/03610927408827101.

Cloutier, L., Pomar, C., Létourneau Montminy, M.P., Bernier, J.F., Pomar, J., 2015. Evaluation of a method estimating real-time individual lysine requirements in two lines of growing-finishing pigs. Animal 9, 561–568. https://doi.org/10.1017/S1751731114003073 http://www.journals.cambridge.org/abstract_S1751731114003073.

Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E., 2008. Querying and mining of time series data. Proceedings of the VLDB Endowment 1, 1542–1552. https://doi.org/10.14778/1454159.1454226 http://dl.acm.org/citation.cfm?doid=1454159.1454226.

Dourmad, J.Y., Etienne, M., Noblet, J., 1991. Contribution à l'étude des besoins en acides aminés de la truie en lactation, in: Journées de la Recherche Porcine, IFIP-Institut du Porc.

Eissen, J.J., Kanis, E., Kemp, B., 2000. Sow factors affecting voluntary feed intake during lactation. Livestock Production Science 64, 147–165. https://doi.org/10.1016/S0301-6226(99)00153-0 http://linkinghub.elsevier.com/retrieve/pii/S0301622699001530.

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

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. J. Anim. Sci. 97, 2822–2836. https://doi.org/10.1093/jas/skz167 https://academic.oup.com/jas/advance-article/doi/10.1093/jas/skz167/5494821 https://academic.oup.com/jas/article/97/7/2822/5494821.

Göransson, L., 1989. The Effect of Feed Allowance in Late Pregnancy on the Occurrence of Agalactia Post Partum in the Sow. J. Vet. Med. Ser. A 36, 505–513. https://doi.org/10.1111/j.1439-0442.1989.tb00760.x http://doi.wiley.com/10.1111/j.1439-0442.1989.tb00760.x.

Koketsu, Y., Dial, G.D., Marsh, W.E., 1994. Association of feed intake patterns in different stages of lactation and reproductive performance, in: 13. International Pig Veterinary Society Congress, Bangkok (Thailand), 26–30 Jun 1994.

Koketsu, Y., Dial, G.D., Pettigrew, J.E., Marsh, W.E., King, V.L., 1996. Characterization of feed intake patterns during lactation in commercial swine herds. Journal of animal science 74, 1202–1210 https://dl.sciencesocieties.org/publications/jas/abstracts/74/6/1202.

Martel, G., 2008. Pratiques d'élevage, productivité des troupeaux de truies et rythmes de travail des éleveurs en production porcine: une approche par modélisation. Ph.D. thesis. http://prodinra.inra.fr/ft?id=%7BAC9CFBB9-77BC-47B7-953C-EE331D588EE7%7D&original=true.

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. https://academic.oup.com/jas/article/68/2/562-572/4631828, doi:10.2527/1990.682562x, arXiv:animres:2002012.

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

O'Grady, J., Lynch, P., Kearney, P., 1985. Voluntary feed intake by lactating sows. Livestock Production Science 12, 355–365. https://doi.org/10.1016/0301-6226(85)90134-4 https://linkinghub.elsevier.com/retrieve/pii/0301622685901344.

Paparrizos, J., Gravano, L., 2016. k-Shape. ACM. SIGMOD Record 45, 69–76. https://doi.org/10.1145/2949741.2949758 http://dl.acm.org/citation.cfm?doid=2949741.2949758.

Piñeiro, C., Morales, J., Rodríguez, M., Aparicio, M., Manzanilla, E.G., Koketsu, Y., 2019. Big (pig) data and the internet of the swine things: a new paradigm in the industry. Animal Frontiers 9, 6–15. https://academic.oup.com/af/article/9/2/6/5448574, doi:10.1093/af/vfz002.

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

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7 https://linkinghub.elsevier.com/retrieve/pii/0377042787901257.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7 https://linkinghub.elsevier.com/retrieve/pii/0377042787901257.

Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust. Speech Signal Process. 26, 43–49. https://doi.org/10.1109/TASSP.1978.1163055 http://ieeexplore.ieee.org/document/1163055/.

Schinckel, A.P., Schwab, C.R., Duttlinger, V.M., Einstein, M.E., 2010. Analyses of Feed and Energy Intakes During Lactation for Three Breeds of Sows. Professional Animal Scientist 26, 35–50. https://doi.org/10.15232/S1080-7446(15)30556-8.

Staicu, A., Islam, M.N., Dumitru, R., van Heugten, E., 2020. Longitudinal dynamic functional regression. J. Roy. Stat. Soc.: Ser. C (Appl. Stat.) 69, 25–46. https://doi.org/10.1111/rssc.12376 https://onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12376.

Theil, P., 2015. 7. Transition feeding of sows, in: Farmer, C. (Ed.), The gestating and lactating sow. Wageningen Academic Publishers, The Netherlands, pp. 147–172. http://www.wageningenacademic.com/doi/10.3920/978-90-8686-803-2_7 https://www.wageningenacademic.com/doi/10.3920/978-90-8686-803-2_7, doi:10.3920/978-90-8686-803-2_7.

Vilas Boas Ribeiro, B.P., Lanferdini, E., Palencia, J.Y.P., Lemes, M.A.G., Teixeira de Abreu, M.L., de Souza Cantarelli, V., Ferreira, R.A., 2018. Heat negatively affects lactating swine: A meta-analysis. Journal of Thermal Biology 74, 325–330. URL https://linkinghub.elsevier.com/retrieve/pii/S030645651830127X, doi:10.1016/j.jtherbio.2018.04.015.

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

Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., Keogh, E., 2013. Experimental comparison of representation methods and distance measures for time series data. Data Min. Knowl. Disc. 26, 275–309. https://doi.org/10.1007/s10618-012-0250-5 http://link.springer.com/10.1007/s10618-012-0250-5.

Warren Liao, T., 2005. Clustering of time series data–a survey. Pattern Recogn. 38, 1857–1874. https://doi.org/10.1016/j.patcog.2005.01.025 https://linkinghub.elsevier.com/retrieve/pii/S0031320305001305.