Selection of methods to analyse body weight and feed intake data used as inputs for nutritional models and precision feeding in pigs

L. Brossard[1], I. Taoussi[2], J. van Milgen[1], J.Y. Dourmad[1]
[1]*PEGASE, INRA-Agrocampus Ouest, 35590 Saint-Gilles, France*
[2]*INSA Rennes, 20 Avenue des Buttes de Coesmes, 35708 Rennes, France*
ludovic.brossard@inra.fr

Abstract

The progress of technologies (sensors, automates) in precision livestock farming enables the development of innovative feeding techniques such as precision feeding of individual animals. In addition to the design of adapted feeders, precision feeding requires decision-support tools to manage data and apply nutritional models that calculate the optimal feed composition and allowance. These calculations require to forecast body weight (BW) and feed intake (FI) of individual pigs according to past performance. To select the most accurate forecasting method, three statistical methods were tested on a dataset of measurements of BW and FI for 117 pigs: the double exponential smoothing (DES) method, multivariate adaptive regression splines (MARS), and the k-nearest neighbours (kNN) method. These methods were tested in relation to data sampling frequency (i.e., daily or weekly measurements) and data availability. The capacity to forecast BW or FI was evaluated through the mean error of prediction. The kNN method appeared suitable if few historical data are available as it requires not more than 3 historical data. The MARS method was better than the DES method to forecast daily BW, but the DES method was better in forecasting the daily cumulated FI. The DES method also seemed more appropriate for weekly BW data, requiring only 3 historical data to make a forecasting. These methods can be used for performance forecasting in a decision-support tool for precision feeding. This study was performed in the Feed-a-Gene Project funded by the European Union's H2020 Programme (grant agreement no 633531).

Keywords: pig, precision feeding, body weight, feed intake, real-time analysis

Introduction

As other livestock production systems, pig production is continuously facing the challenge of sustainability. To contribute to the growing demand for animal protein, feed efficiency in pig production has to be optimised. This optimisation will contribute to reduce the environmental impact and to improve competitiveness of a production system where feed represents a major part of the production costs (typically 60 to 70%). The progress of technologies (sensors, automates) in precision livestock farming enables the development of potentially novel feeding techniques such as precision feeding. Precision feeding is based on the dynamic adjustment (if possible day by day) of the nutrient supply to the requirement, at a group or at an individual level (Pomar *et al.*, 2009). Recent studies have shown that precision feeding is a promising way to improve feed efficiency (e.g., Andretta *et al.*, 2014). Compared to common phase-feeding programs applied to groups of pig, precision feeding allows a better consideration of the change in requirements during growth and of the variability among pigs of the same age, sex, and genotype (Brossard et al., 2009).

In addition to the design of adapted feeders (e.g., Pomar *et al.* (2009) for *ad libitum* feeding, Marcon *et al.* (2015) for restricted feeding), and a precise evaluation of feed ingredients, precision feeding requires decision-support tools to manage data and apply nutritional models that calculate the optimal feed composition and allowance. Current models developed to simulate pig growth and to determine nutrient requirements (e.g. van Milgen *et al.*, 2008) are difficult to be used in real-time decision-support tools as they require historical data on body weight (BW) and daily feed intake (DFI) data to characterize the animal. However, in precision feeding, nutrient requirements for individuals need to be determined in real time on the basis of their own growth and FI patterns. Hauschild *et al.* (2012) proposed a model for pigs fed *ad libitum* with empirical and mechanistic components, where the empirical model allowed the estimation of DFI and BW at time t+i from historical information measured for each individual animal up to time t. Hauschild *et al.* (2012) used the double exponential smoothing (DES) forecasting time series method on DFI and on weekly BW. However, with the technological development of sensors, BW can now also be obtained daily and individually (e.g., Marcon *et al.*, 2015).

The aim of this study was to test the most accurate forecasting method for BW and DFI for *ad libitum* feed allowance among three statistical methods: the DES method, the multivariate adaptive regression splines (MARS) method, and the k-nearest neighbours (kNN) method. These methods were tested in relation to data sampling frequency (i.e., daily or weekly measurements) and data availability.

Material and Methods

Dataset

A dataset on 119 pigs from Topigs was used for the analyses and calculations. During collection of data, animals had *ad libitum* access successively to two diets formulated to meet or exceed nutritional requirements (National Research Council, 1998), with a diet change at 65 kg BW: a growing diet (18.7% crude protein (CP), 10.51 MJ net energy (NE)/kg, 1.06 g standard ileal digestible lysine (SID Lys) / MJ NE on an as-fed basis), and a finishing diet (15.6% CP, 10.24 MJ NE/kg and 0.89 g SID Lys / MJ NE on an as-fed basis). During the test, DFI was recorded for each pig using automatic feeder systems. Animals were weighed daily by automatic weighing devices. Mean BW at the beginning and end of the data collection period were 34.0 ± 4.7 kg (75.9 ± 6.6 d of age) and 139.9 ± 9.8 kg (176.0 ± 9.2 d of age), respectively. Average DFI observed during the data collection was 2.32 ± 0.73 kg/d. On average, 99 and 101 data were available per animal for BW and DFI, respectively.

Calculation methods

To estimate the BW or DFI at a time t+i, using data up to time t, three forecasting methods were tested: the DES method, the MARS method, and the kNN method.

Double exponential smoothing (DES) method

In the panel forecasting time series methods, exponential smooting techniques are appropriate to reduce fluctuations from irregular observations in the studied time series (Claycombe and Sullivan, 1977), with a goal of short-time forecasting. As indicated by Hauschild *et al.* (2012), the DES is well adapted to the study of DFI and BW in pigs as they show long-run trends without seasonal components and because the method works with a limited number of observations (at least 3). At a given time t, the DES forecasting of the series to i days ahead is given by:

$$\hat{X}_t(i) = a_t i + b_t$$

where the coefficients $\mathbf{a_t}$ and $\mathbf{b_t}$ vary with time. In the DES method, the smoothed series $\mathbf{S_1}$ is smoothed a second time to obtain $\mathbf{S_2}$ with

$$S_1(t) = \alpha y_t + (1-\alpha)\, S_1(t-1) \quad and \quad S_2(t) = \alpha S_1(t) + (1-\alpha) S_2(t-1)$$

where α is the smoothing constant comprised between 0 and 1 and used to weigh differently past and recent observations. Values of α close to 1 give a greater weight to recent observations, while values of α close to 0 have a greater smoothing effect and are less responsive to recent changes.

Multivariate adaptive regression splines (MARS) method

The MARS method is an adaptation of techniques developed by Friedman (1991) to resolve regression problems, with the aim to predict values of one or several continuous variables using a set of explicative variables. This non-parametric method has been largely used in data mining because it does not require a hypothesis on residues or relationships. The MARS method can be seen as an extension of linear regression to model automatically interactions and non-linearity. The method uses a database to establish functional relationships between explicative and predicted variables, even if the relationship between variables is not continuous and difficult to approximate with parametric models. In the MARS method, these relationships are approximate using linear-based functions such as:

$$B(x) = \begin{cases} (x - t) \ \ if \ x \geq t \\ \quad else \ 0 \end{cases}$$

where t points are nodes connecting two regression segments. The general equation of the model for the variable t depending on the variable x is obtained by combining the linear-based functions estimated through the least squares method:

$$y = f(t) = \beta_0 + \sum_{m=0}^{M} \beta_m B_m(t)$$

where B_m are linear-based functions with associated coefficients β_m.

K-nearest neighbours (kNN) method

The kNN method is an intuitive and non-parametric method used for classification and regression (Altman, 1992). The method is based on the determination of the k-nearest neighbours in a population of training values. In the regression application, the predicted value is the average of the k-nearest neighbours. To forecast a variable value at time t, the kNN method requires the use of historical data from the individual to be forecasted, and a learning database with values for previous days and time t for a population of individuals. Different parameters have to be determined for the calculation: the number k of neighbours, the type of distance to be calculated between individuals, the possible weighing of distance in calculation of the average (to assign weights to the contributions of the neighbours, so that the nearer neighbours contribute more to the average than the more distant ones), the number of previous days to be used to determine the nearest neighbours, and the weight to give to recent observations compared to past observations. For continuous variables such as BW or DFI, the Euclidian distance is commonly used. Preliminary tests we made showed that a triweight kernel and k = 3 could be considered for calculations on BW and DFI.

Calculations

For daily measurements of FI and BW, the forecasting performance at day+1 of the DES and MARS methods were compared. For the DES method, values for the smoothing constant α ranging from 0.1 to 0.9 by a 0.1 increment were tested. As DFI can vary considerably from one day to another, we considered that the forecasting of DFI may be too sensitive to forecast these variations. Consequently, and similarly to BW, the cumulative FI was forecasted. The DFI can be determined from the cumulative FI.

For weekly available data, we considered that forecasting of DFI is required for the application of precision feeding. Consequently only forecasting of BW at day+7 was tested. The MARS method requires at least 8 historical data to perform forecasting whereas the DES method requires 3 data. Considering that this limits the usefulness of the MARS method, only the DES method was tested on BW for weekly data, with α ranging from 0.1 to 0.9. BW data at 7 day intervals were extracted from the original dataset to perform calculations.

The kNN method was used to illustrate the possibility to forecast BW or DFI at day+1 when only 1 or 2 historical data are available. This occurs for example at the beginning of data collection where the MARS or DES methods cannot be used. In the original dataset, data of 83 animals were used to create a learning database to forecast BW and DFI of the 36 other pigs.

All calculations were performed every day (or week) for each pig using the R software (version 3.3.2). The following functions and R packages were used: the earth function from the earth package (Milborrow, 2011) for the MARS method; the HoltWinters function from the stats package for the DES method; the kkNN function from the FNN (Beygelzimer *et al.*, 2013) and kknn (Schliep and Hechenbichler, 2016) packages.

Number of previous data used in calculation

The DES and MARS methods were tested on daily BW and cumulative DFI data using the 8, 13 or 20 historical (i.e., latest) data (the 8 data corresponds to the minimal number of data for the MARS method). Moreover, the 8 historical data refer approximatively to one week of data recording. The 13 and 20 historical data were chosen as they refer to a BW or DFI forecasting at 14th or 21st days, i.e. at the end of the 2 or 3 last weeks of data collection. Thus, the calculation started at day 9 and was performed on 8 historical data or integrated progressively up to 13 or 20 historical data depending on the targeted number of historical data. For the weekly BW, the calculation started at day 4 and a maximum of 8 historical data was used.

Missing data

For some pigs, data for BW or DFI were missing in the dataset. As the methods used cannot deal with missing values, missing data were created to obtain a specific completed datasets for each tested method. For data before day 4, BW and DFI were created by adding 0.75 kg and 27 g to previous BW and DFI, respectively. From day 4, data were created using the tested method with corresponding number of historical data.

Statistics

The residual mean square error of prediction (RMSEP) was calculated between forecasted values and measured value for each pig, excluding the first 8 days for tests on DFI and daily BW forecasting, and the first 3 days for weekly BW forescasting. The RMSEP were submitted to an analysis of variance (proc MIXED, SAS v9.4, Inst. Inc. Cary, NC). Least-square means were compared. For the daily BW or cumulative FI, the main effects were the forecasting method (with MARS and DES with α ranging from 0.1 to 0.9), the number of historical data (8, 13 or 20) and their interaction. For the weekly BW, the main effect was the α value (ranging from 0.1 to 0.9).

Results and Discussion

Daily forecasting of BW

The RMSEP of daily BW decreased with increasing number of historical data used (Table 1). However, this decrease was significant only from 8 to 13 data for MARS method and α = 0.1 and 0.3 for DES method and from 13 to 20 data for α = 0.3 (method x number of historical data interaction, $P < 0.001$). The lowest RMSEP (1.1 kg) were obtained with the MARS method (13 or 20 historical data). This corresponds to 3% and 0.7% of BW at the beginning and the end of test period, respectively. For the DES method, the lowest RMSEP were obtained with α = 0.3 to 0.6. These α values give an intermediate weight between recent and less recent data. This means that most recent or oldest historical data should not be given too much weight to forecast BW. The results indicate that the MARS method used with 13 to 20 historical data could be preferred to the DES method for daily forecasting of BW, avoiding the choice of an α value. Quiniou et al. (2017) observed that the DES method with α = 0.6 to obtain the best RMSEP for data from pigs restrictively fed. Hauschild et al. (2012) used α = 0.1 to forecast BW (BW range = 25 to 105 kg, without reporting the RMSEP), giving a high weight to less recent data. This indicates that such comparisons could be influenced in part by the BW range or by the feeding level that can affect BW.

Weekly forecasting of BW

The RMSEP of weekly BW were lowest with α = 0.5 to 0.6 (2.14 kg; Table 2) and increased for lower or higher α values. The highest RMSEP were obtained with α = 0.1 and 0.2. As for daily BW forecasting, giving a quite balanced weight to recent and less recent data allowed to better forecast BW, especially compared to privileging oldest data. The RMSEP values observed for weekly BW were 1 to more than 2 kg higher than for daily BW forecasting. The BW varies from day to day because of growth but also due to eating, defecating and urinating patterns. Consequently the precision of forecasting is affected when using weekly data where the difference between two successive points can be sensitive to the conditions of BW measurement.

Table 1: RMSEP (kg) of daily BW forecasting using the double exponential smoothing (α value ranging from 0.1 to 0.9) or MARS methods and 8, 13, or 20 historical data[1].

Nb. of data	Method									
	MARS	Double exponential smoothing (α value)								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
8	1.34[a]	2.00[b]	1.23[c]	1.29[ac]	1.21[cd]	1.19[cd]	1.22[c]	1.27[ac]	1.34[a]	1.46[e]
13	1.13[a]	1.79[b]	1.27[c]	1.21[cd]	1.18[ad]	1.19[ac]	1.22[cd]	1.27[c]	1.34[c]	1.46[e]
20	1.11[a]	1.43[b]	1.26[c]	1.19[cd]	1.18[ad]	1.19[cd]	1.22[cd]	1.27[ce]	1.34[e]	1.46[b]

1. Least-square means. The main effects of method, number of historical data and their interaction were significant at P < 0.001 (residual standard deviation of the model = 0.31 kg). Within a row, values followed by common letters are not significantly different for the method effect (P < 0.05).

Table 2: RMSEP (kg) of weekly BW forecasting using the double exponential smoothing method on 8 historical data with α values ranging from 0.1 to 0.9[1].

	α value								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
RMSEP (kg)	4.18[a]	2.79[b]	2.43[c]	2.23[cd]	2.14[d]	2.14[d]	2.19[d]	2.28[cde]	2.45[cde]

1. Least-square means. Effect of α value significant at P < 0.001 (residual standard deviation = 0.90 kg). Values followed by common letters are not significantly different (P < 0.05).

Daily forecasting of cumulated FI

With the MARS method, the RMSEP of cumulated FI significantly increased with an increasing number of historical data used (Table 3). With the DES method, the number of historical data did not influence the RMSEP (P > 0.05), except for α = 0.2 where the RMSEP was significantly lower for 8 historical data compared to 13 or 20 historical data. The lowest RMSEP values (0.49 kg) were obtained with the DES method and α = 0.6 to 0.8, even though the difference between the MARS and DES methods was not significant for 8 historical data. These RMSEP values are lower than those obtained for daily BW, despite the fact that the daily increase in FI is higher than in BW (ADFI = 2.3 kg/d vs average daily gain = 1.06 kg/d) and the variation in DFI can be important from a day to another. This potential less smooth evolution of cumulative FI could also explain the increase in RMSEP with increasing number of historical data observed with the MARS method that uses combination of linear regressions. For the cumulative FI, the best forecasting was obtained by giving a higher weight to more recent data (α > 0.6). In contrast, Hauschild et al. (2012) used α = 0.1 to forecast DFI (rather than cumulated FI). The present results indicate that the DES method used with 8 to 20 historical data could be preferred to the MARS method for daily forecasting of the cumulative FI, with α = 0.6 to 0.8, taking advantage of using a larger number of available historical data (more than 8) with a better RMSEP.

Table 3: RMSEP (kg) of cumulated FI forecasting using the double exponential smoothing (α value ranging from 0.1 to 0.9) or MARS methods and 8, 13, or 20 historical data[1].

Nb. of data	Method									
	MARS	Double exponential smoothing (α value)								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
8	0.52^{a}	1.06^{b}	0.67^{c}	0.60^{d}	0.54^{ae}	0.50^{ae}	0.48^{a}	0.48^{a}	0.49^{a}	0.51^{a}
13	0.60^{a}	1.04^{b}	0.73^{c}	0.61^{a}	0.54^{d}	0.50^{de}	0.49^{e}	0.49^{e}	0.49^{e}	0.52^{de}
20	0.70^{a}	1.02^{b}	0.75^{c}	0.61^{e}	0.54^{f}	0.51^{fg}	0.49^{g}	0.49^{g}	0.50^{fg}	0.52^{fg}

1. Least-square means. The main effects of method, number of historical data and their interaction were significant at P < 0.001 (residual standard deviation of the model = 0.17 kg). Within a row, values followed by common letters are not significantly different for the method effect (P < 0.05).

The kNN method for obtaining initial data

We forecasted the BW and DFI with the kNN method at day 2 and 3 of data collection, on the basis of the first day or first 2 days of available data. The RMSEP of BW and DFI were 0.86 kg (± 0.73 kg) (n = 35 due to an outlier) and 0.33 kg (± 0.16 kg) (n = 36), respectively. This RMSEP was lower than for the two other methods although obtained with fewer data. The RMSEP of BW was higher and more variable than that for DFI, probably due to differences in the absolute value between BW and DFI at this stage (38 kg vs 1.27 kg/d, respectively). This method can be useful for forecasting when other methods cannot be used. However it requires a database on BW and DFI obtained in similar pigs reared in similar conditions. It is likely to be more sensitive to day-to-day variation than the DES or MARS methods that can smooth variation by using a larger number of historical data for the same pig.

Conclusions

The results of this study indicated that the MARS method used with 13 to 20 historical data is to be preferred to the DES method for daily forecasting of BW. Conversely the DES method is preferred to the MARS method for daily forecasting of cumulative FI, and can be used to forecast weekly BW. The kNN method can be useful to forecast BW or DFI at the start of data collection, when the two other methods cannot be used. The results about DES and MARS methods have to be confirmed on larger datasets and on different rearing conditions (e.g., with restricted feeding). However, the methods can be integrated for an efficient forecasting of BW and FI in a decision-support tool for precision feeding.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant N° 633531. The authors gratefully acknowledge N. Quiniou and M. Marcon (IFIP) for methodological support.

References

Altman, N. S. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* **46** 175–185.

Andretta, I., Pomar, C., Rivest, J., Pomar, J., Lovatto, P.A., and Neto, J.R. 2014. The impact of feeding growing-fisnihing pigs with daily tailored diets using precision feeding techniques on animal performance, nutrient utilization, and body and carcass composition. *Journal of Animal Science* **92** 3925-3936.

Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., and Li, S. 2013. *FNN: Fast Nearest Neighbor Search Algorithms and Applications*. R package version 1.1. https://CRAN.R-project.org/package=FNN

Brossard, L., Dourmad, J.-Y., Rivest, J., and van Milgen, J. 2009. Modelling the variation in performance of a population of growing pig as affected by lysine supply and feeding strategy. *Animal* **3** 1114-–1123.

Claycombe, W., and Sullivan, W. 1977. *Fundamentals of forecasting*. Reston Publishing, Reston, VA.

Friedman, J. H. 1991. Multivariate Adaptive Regression Splines. *Annals of Statistics* **19** 1–67.

Hauschild, L., Lovatto, P. A., Pomar, J., and Pomar, C. 2012, Development of sustainable precision farming systems for swine: Estimating real-time individual amino acid requirements in growing-finishing pigs. *Journal of Animal Science* **90** 2255-–2263.

Marcon, M., Brossard, L., and Quiniou, N. 2015. Precision feeding based on individual daily body weight of pigs with an automatic feeder developed to allow for restricting feed allowance. In: *Proc. of the 7th European conference on Precision Livestock Farming*, Guarine M. and Berckmans D. (Ed), Milan, Italy, 592-602.

Milborrow, S. 2011. *Earth: multivariate adaptive regression spline models.* R package version 4.4.9.1. URL http://CRAN.R-project.org/package=earth

National Research Council (NRC), 1998. *Nutrient requirements of swine, 10th revised edition.* National Academic Press, Washington, DC, USA.

Pomar, C., Hauschildt, L., Zhang, G.H., Pomar, J., and Lovatto, P.A. 2009. Applying precision feeding techniques in growing-finsihing pig operations. *Revista Brasileira de Zootechnia* **38** 226-237 (supl. especial).

Quiniou, N., Brossard, L., and Marcon, M. 2017. Assessment of the dynamic growth of the fattening pigs from body weight measured daily and automatically to elaborate precision feeding strategies In: *Proc. of the 8th European conference on Precision Livestock Farming*, XXXX and XXX (Ed), Nantes, France, XX-XX.

R Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Schliep, S., and Hechenbichler, K. 2016. *kknn: Weighted k-Nearest Neighbors.* R package version 1.3.1. https://CRAN.R-project.org/package=kknn

van Milgen, J., Valancogne, A., Dubois, S., Dourmad, J. Y., Sève, B., and Noblet, J. 2008. InraPorc: A model and decision support tool for the nutrition of growing pigs. *Animal Feed Science and Technology* **143** 387-405.