Computers and Electronics in Agriculture 128 (2016) 100–111

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

An automatic model configuration and optimization system for milk production forecasting

Fan Zhang [a], Michael D. Murphy [a,*], Laurence Shalloo [b], Elodie Ruelle [b], John Upton [b]

[a] *Department of Process, Energy and Transport Engineering, Cork Institute of Technology, Co. Cork, Ireland*
[b] *Animal and Grassland Research Innovation Centre, Teagasc Moorepark, Co. Cork, Ireland*

ARTICLE INFO

Article history:
Received 12 April 2016
Received in revised form 11 August 2016
Accepted 20 August 2016
Available online 1 September 2016

Keywords:
Dairy production
Milk production forecasting
Dairy modelling

ABSTRACT

The objective of this study was to develop and implement the Milk Production Forecast Optimization System (MPFOS) for the purpose of comparing the effectiveness of multiple herd milk yield prediction models for an Irish pasture-based dairy herd. The MPFOS was populated by nine milk production models that were categorized into three types: curve fitting, regression and auto-regressive models. The Adaptive Stratified Sampling Approach (ASSA) was introduced for data filtering, processing and for randomly selecting each member of the 100 cow sample herd. The MPFOS calculated optimal model parameters, statistical analysis and milk production forecasts for each chosen model using input data combinations based on animal, herd and milk production records. The model evaluations were based on historical milk production data between the years 2004 and 2009 from dairy farms in the south of Ireland situated in close proximity. Milk yield records from 2004 to 2008 were used for model training, whereas the milk production records for 2009 were set for model evaluation and validation. The ASSA randomly selected the representative herd population based on the required criteria. The MPFOS automatically generated the optimal configuration for each of the nine milk production forecast models and benchmarked their performance over a short, medium and long term prediction horizon. The Root Mean Square Error (RMSE) value of the nine prediction models varied substantially (from 68.5 kg to 210.4 kg per day). The surface fitting model performed better (10% in RPE and R^2) than the dynamic NARX model for the same prediction horizon (365-day and 30-day). However, the NARX model provided more accurate results for shorter (10-day) prediction horizons. The MPFOS found the most accurate model based on prediction horizon length and on number of input parameters. The results of this study demonstrate the effectiveness of the MPFOS as a model configuration and comparison tool. The MPFOS may also be employed for selecting the optimal milk production forecast model for a specific application.

1. Introduction

Milking quotas were abolished in the European Union in April 2015. In an environment unconstrained by milk quotas, milk delivery projections will become increasingly important at both farm and processor levels for cash flow planning, feed budgeting, marketing and planning future adjustments in processing capacity. Accurate milk production forecasts would allow farmers to predict on farm thermal cooling loads, plant capacity sizing, and to optimize plant configurations as well as cash flow planning. Concurrently, accurate milk production forecasts will be useful for farm management support and analysis for herd management, energy utilization and economic prediction (Shalloo et al., 2004, 2011; Murphy et al., 2013; Upton et al., 2015).

Historically, studies have been undertaken regarding milk production prediction techniques where diverse equations have been developed for the purpose of describing a lactation curve based on past milk yield data, including: curve fitting models, regression models and auto-regressive models. Curve fitting models usually require one single variable as input data, such as daily or weekly cumulative milk yield of a herd or of an individual cow. Curve fitting category models have proven to perform well by many authors in specific studies with many different forms including parabolic

Abbreviations: MPFOS, milk production forecast optimization system; ASSA, adaptive stratified sampling approach; DIM, days in milk; NCM, number of cows milked; DHMY, daily herd milk yield; SANN, static artificial neural networks; ANN, artificial neural network; MLR, multiple linear regression; NARX, nonlinear auto regressive model with exogenous input; SSE, summed square of residual errors; R-square, coefficient of determination; RMSE, root mean square error; RPE, relative prediction error; GUI, general user interface.

* Corresponding author.
E-mail address: MichaelD.Murphy@cit.ie (M.D. Murphy).

http://dx.doi.org/10.1016/j.compag.2016.08.016
0168-1699/

exponential (Sikka, 1950), incomplete gamma (Wood, 1967), polynomial (Ali and Schaeffer, 1987), exponential (Wilmink, 1987), cubic splines (Green and Silverman, 1993), Legendre polynomial (Kirkpatrick et al., 1994) and log-quadratic (Adediran et al., 2012). However, lack of flexibility and adaptation is a common weakness of curve fitting models when dealing with significant fluctuations in yield within and between years (Jones, 1997). Static artificial neural networks (SANN) (Lacroix et al., 1995; Salehi et al., 1998; Sharma et al., 2007) and conventional multiple linear regression (MLR) models (Sharma and Kasana, 2006) are regression category models that have been found to provide accurate predictions with varied amounts of detailed input data from basic daily herd yield to complicated individual traits such as genetic group, period of birth, peak yield and weight at calving. The non-linear auto regressive with exogenous input (NARX) model (Murphy et al., 2014) is an auto-regressive category model that has shown to produce more accurate milk production predictions when compared to the MLR model and the SANN models, especially in the short term due to its short term embedded memory and ability to dynamically adapt its prediction trajectory. Similar studies have been carried out regarding the comparison of different modelling techniques within the same category (Adediran et al., 2012; Quinn et al., 2005; Silvestre et al., 2006) and cross-category (Grzesiak et al., 2006; Murphy et al., 2014). Especially for comparison within two categories, numerous works have been carried out across different model development and results evaluation platforms (Adediran et al., 2012; Cole and VanRaden, 2006; Grzesiak et al., 2006; Sharma and Kasana, 2006; Van Bebber et al., 1999) whereby these configurations may increase the complexity and time consumption of the prediction model. Furthermore, cross category milk yield model comparisons are technically and computationally more complex than those within the same category. There is a requirement for the forecasting of herd milk yield from an information integrated perspective where the solution can integrate full features including data gathering, storage and processing, all categories model configuration, simulation and optimization, results analysis and optimal prediction calculation.

The aim of this study was to develop and demonstrate the Milk Production Forecast Optimization System (MPFOS) with the Adaptive Stratified Sampling Approach (ASSA) for automatic model configuration, comparison, optimization and validation. The MPFOS architecture was designed to calculate model parameters for curve fitting techniques, to calculate coefficients for regression models, to select optimal training algorithms and neuron architectures for neural network models and duration for auto-regressive memory. The ASSA filters and sorts the input data to ensure the training dataset is representative of the entire population. The final output of the MPFOS contains configurations for each prediction model, statistical analysis for all simulation results and the optimal milk production forecast. In short, the MPFOS selects the most effective milk production forecast model and corresponding model configuration for a specific cow population. While numerous model categories and model configurations have been found to be most effective for a particular dairy cow group in previous studies, no one model has shown to produce the most accurate milk production forecast for all circumstances. The results in Section 4 demonstrate the capability and performance of the MPFOS.

2. The MPFOS architecture

2.1. Design of the MPFOS architecture

The MPFOS focuses on global data processing, automated model configuration and optimization and can accomplish multiple model comparisons at a global level. The self-adaptive capability of the MPFOS can provide automatic configurations for different modelling techniques by providing corresponding input datasets. Once various well-known models were translated into algorithms, implemented as programming code and stored in the MPFOS as specific files, all possible subsequent repetitive work is avoided, with the modelling techniques abstracted, thus requiring no further manual interventions from the user side. MPFOS can calculate parameters, coefficients or optimal training configurations for corresponding category models automatically with the same input training dataset in one multiple model comparison procedure. More importantly, all relevant data for simulation and calculation are stored in databases of the MPFOS which can be reused for future data analysis. The space requirements for the empirical data vary as different category milk yield prediction models require various input data combinations and hence corresponding output results have differing degrees of accuracy.

Three different categories of milk yield prediction models were chosen in the model library of the MPFOS including curve fitting models, regression models and auto-regressive models. The primary reason for choosing these three model types is that in consideration of other authors' studies and conclusions, each one of these models has been successfully applied to cow/herd level milk production modelling, based on specific datasets. For example, the adaptive polynomial model (Quinn et al., 2005) was the best fitting model for Irish experimental study data in 2005, the log-quadratic model (Adediran et al., 2012) was optimal and recommended for Australian pasture-based data in 2012, an artificial neural network (ANN) model was superior to the conventional MLR model (Sharma et al., 2007) and in 2014, the nonlinear auto-regressive (NARX) model was introduced for milk yield prediction and presented more accurate forecasting compared with the ANN model (Murphy et al., 2014). Therefore, nine representative models of three categories were chosen to populate the MPFOS (detailed information regarding formulae for these nine milk yield prediction models is available within the Appendix A).

The primary challenge in carrying out a model comparison between two or more model categories is that different models have unique data input formatting requirements. It should be emphasized that the optimal model was dependent on the training input dataset in many cases. For example, daily herd milk yield (DHMY) and corresponding days in milk (DIM) are essential and common training inputs for all milk prediction models and especially for curve fitting category models. Besides DHMY and DIM, number of cows milked (NCM) was selected as a data input for regression and surface fitting models. Additional input information such as calving date (McCarthy et al., 2013), parity and meteorological conditions could be incorporated into the ANN and NARX models. It is reasonable to extend the scope of model comparisons to test as many milk yield prediction models and input combinations as possible. Therefore, the MPFOS has the ability to comprehensively simulate each populated milk prediction model with all possible combinations of input data, compare the accuracy of every scenario and calculate the optimal model configuration.

Fig. 1 shows the database design inside the MPFOS with a brief description. Three separated databases exist for functionality and scalability in the MPFOS, including the milk yield database, the cow description database and the weather database. With the possibility of performing more experiments for future hypothesis, the architecture was designed to allow the database to be extended through a greater number of training data inputs, such as milk composition records including protein and fat content, feeding records and so on.

2.1.1. The milk yield database

The milk yield database contained information related to every daily milking yield record for each cow in the entire dataset.