ISSN 2004-2965 Energy Proceedings, Vol. 21, 2021

International Conference on Applied Energy 2021
Nov. 29 - Dec. 5, 2021, Thailand/Virtual
Paper ID: 027

Predicting milk cooling, milk harvesting and water heating electricity use on dairy farms using artificial neural networks

P. Shine[1*], M. D. Murphy[1]

1 Department of Process, Energy and Transport Engineering, Munster Technological University, Cork, Ireland

ABSTRACT

Artificial neural network (ANN) models were developed to predict milk cooling, milk harvesting and water heating electricity consumption using data collected from 56 pasture-based Irish dairy farms. The methodology employed variable selection, outlier detection, hyper-parameter tuning and nested cross-validation. The ANN models were developed to predict monthly electricity use, while monthly predicted values were also aggregated and assessed at farm- and catchment-levels. Model input variables were constrained to stock and milk production, infrastructural equipment and farm management details. The ANN algorithm predicted monthly electricity consumption for milk harvesting with an error of 22% (relative prediction error), milk cooling to within 24% and water heating to within 31%. Prediction errors reduced to 16%, 12% and 9%, respectively when predicted values were aggregated at the farm-level. In addition, significant reductions in prediction errors were calculated when *m*ilk harvesting (0.8%), milk cooling (1.8%), and water heating (1.9%) predictions were aggregated at the catchment-level. This demonstrates the potential effectiveness of the developed ANN models as tools for macro-level simulations.

Keywords: energy, modelling, neural network, machine learning, dairy, agricultural sustainability

1. INTRODUCTION

Animal agriculture is responsible for 14.5% of global, human induced greenhouse gas (GHG) emissions, of which 20% is due to cattle milk production [1]. With global milk production forecasted to increase by 22% between 2018 and 2027 [2], it is essential that the dairy sector addresses the significant challenges ahead related to minimizing GHG emissions across the entire production cycle. Although energy consumption is responsible for only 2% of global milk production related GHG emissions across the entire supply chain [1], significant financial savings may be made on-farm through the optimal sizing and operation of energy infrastructural equipment, while also providing advantaged to the electrical grid.

Energy is consumed on dairy farms both directly through electricity and tractor fuel use, and indirectly through the production and delivery of fertilizers, machinery, and concentrate feed, etc. [3]. Electrical energy is responsible for 14% of total primary energy consumed on conventional dairy farms, mainly due to milk cooling (31%), milk harvesting (29%) and water heating (19%) [3].

Machine-learning models have been developed to allow researchers gain a greater understanding of the various farm parameters affecting farm productivity [4], including on-farm electricity use. Multiple linear regression (MLR) models have been developed to predict total electricity consumption (kWh) [5,6], the output energy of milk (MJ Cow^{-1}) [7] and diesel use (kg) [6]. In addition, MLR has been employed to estimate milk cooling [8,9], water heating, milk harvesting and air compressor electricity use [9]. More recently, machine-learning algorithms have been employed to quantify non-linearities and interactions between input variables to increase prediction accuracy. Shine et al. [10] developed support vector machine, decision tree, random forest and artificial neural network (ANN) models to estimate on-farm electricity consumption (kWh). Concurrently, the ANN [11] and adaptive neural-fuzzy inference system [7] algorithms have been employed to predict the output energy of milk (MJ Cow^{-1}) on dairy farms in Iran. However, no study to date has focused on applying machine-learning algorithms to predict milk cooling, milk harvesting and water heating electricity use.

The objectives of the work presented in this paper were to: 1) identify the farm variables that minimize prediction error when estimating milk cooling, milk harvesting and water heating electricity use. 2) Calculate the accuracy of the ANN algorithm when estimating monthly, farm-level and catchment-level milk cooling, milk harvesting and water heating electricity consumption. The methodology employed a range of data mining techniques including: variable selection methods to extract high predictive yielding variables, grid-search hyper-parameter tuning to improve the prediction performance of each ANN model, and stratified nested cross-validation to calculate the prediction performance.

2. MATERIALS AND METHODS

2.1 Data Collection

Data were acquired via the automated recording of electricity consumption on 56 pasture-based dairy farms, located in the south of Ireland. Milk cooling, milk harvesting and water heating electricity consumption data were monitored between 1st Jan 2014 – 30th June 2017. A once off survey was completed on each study farm in 2014 to identify equipment and managerial processes utilized on-farm. Milk yield data was also attained from each farm's milk processor, and stock data was attained from the Irish Cattle Breeding Federation.

2.2 Data pre-processing

Table 1 List of dairy farm variables for model development

Variable Category	Milk Cooling	Milk Harvesting	Water Heating
1. Month number	✓	✓	✓
2. No. lactating cows	✓	✓	✓
3. Total no. cows	✓	✓	✓
4. Milk production	✓	✓	✓
5. No. parlour units	✓	✓	--
6. Milk cooling system	✓	--	--
7. Bulk tank volume	✓	--	--
8. Vacuum pump power	--	✓	--
9. Variable speed drive	--	✓	--
10. Freq. of hot wash	--	--	✓
11. Solar thermal system	--	--	✓
12. Water tank volume	--	--	✓
13. Water heating power	--	--	✓
14. Water heating fuel	--	--	✓

In total, 14 variables were considered to predict milk cooling, milk harvesting and water heating electricity consumption, as shown in Table 1. These variables were related to stock and milk production, infrastructural equipment and farm management. The month number was a fixed input for ANN development to address the time series component of the model. The milk cooling category included: i) whether an ice bank of direct expansion bulk tank was used, ii) whether milk pre-cooling was carried out using ground water, and iii) whether ice cold water was used to pre-cool milk. Similarly, four hot washing frequencies were used throughout study farms: 1) once a day, 2) every second day, 3) once a week, and 4) once a month. Lastly, the water heating fuel source category include either electric water heating or a mixture of electric and oil.

Data were normalized to a mean value of zero and standard deviation of one. All variables then underwent a selection process before model development to reduce the number of possible dimensions and required computational resources and noise within the datasets. Backward and forward sequential variable selection with support vector machine and decision tree models were employed, as described in Shine et al. [10]. Outliers were detected and removed using the Density Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm with a parameter calculation methodology described in Shine et al. [5].

2.3 ANN Development

Three ANN models (multilayer perceptron) were developed to predict monthly milk cooling, milk harvesting and water heating electricity use. The Levenburg Marquardt algorithm with Bayesian regularization [12] was employed to iteratively update synapse weights to reduce the residual sum squared error of the prediction values until no improvement was achieved or once 250 iterations were carried out. Bayesian regularization minimizes a combination of the squared errors and synapse weights, and then selects the optimum combination to produce a network that generalizes well.

To select the optimum number of neurons in the hidden layer, five hidden layer sizes were determined from the number of variable inputs, and assessed via grid-search hyper-parameter tuning. Where I is the number of input variables, the range of hidden layer sizes assessed was [$I-2$, $I-1$, I, $I+1$, $I+2$]. The tan-sigmoid activation function was selected due to previously being found to minimize prediction error for total electricity consumption on dairy farms [10].

2.4 Validation and Performance

The prediction performance of each variable subset selected through each of the four sequential variable selection methods were calculated through a stratified nested cross-validation [10]. The nested cross-validation method utilized 10-fold cross validation in an outer loop and 9-fold cross-validation in an inner loop. The ANN models were developed using the outer loop training data selecting the number of hidden neurons that minimized relative prediction error (RPE) through 9-fold cross-validation. Model accuracy was then calculated on the remaining outer loop test fold, and process repeated until the prediction accuracy was calculated on each 10 outer loop folds. Overall accuracy then equaled the mean across all 10 outer loop test folds.

The ANN models were developed to predict monthly electricity consumption and predicted values were then aggregated (summed) and assessed at farm- and catchment-levels. Prediction biases were evaluated according to the mean percentage error (MPE (%)), with negative MPE values suggesting overprediction, and positive values suggesting underprediction. Absolute model precision was evaluated according to relative prediction error (RPE (%)), root mean squared error (RMSE (kWh)), and the concordance correlation coefficient (CCC), as described in Shine et al. [10].

3. RESULTS

3.1 Monthly Prediction Accuracy

Table 2 Monthly prediction accuracy of each ANN model

Target	V/S*	RPE	CCC	RMSE
Milk Harvesting	BW - SVM	22%	0.89	93 kWh
Milk Cooling	BW - SVM	24%	0.94	153 kWh
Water Heating	FW - DT	31%	0.90	125 kWh

*V/S = Variable selection method; BW - SVR = Backward variable selection with support vector machine; FW - DT = Forward variable selection with decision tree

The ANN algorithm predicted monthly milk harvesting electricity use with an RPE of 22% (n = 1,452), milk cooling to within 24% (n = 1,514) and water heating to within 31% (n = 1,413), as shown in Table 2. The total number of dairy cows, milk yield (L), the number of parlor units and vacuum pump power (kW) offered the minimum prediction error of milk harvesting electricity use. Regarding milk cooling, the total number of dairy cows, milk yield (L), the number of parlor units, the type of milk cooling system used (type of bulk tank (ice bank or direct expansion), whether milk pre-cooling was carried out with or without ground water (Yes | No) or ice-cold water (Yes | No)), and the total bulk tank volume (L). Regarding water heating, the number of lactating cows, the total number of dairy cows, the frequency of hot washing (whether carried out once a day, every second day or once a week), whether a solar thermal system was utilized to heat hot water (Yes | No), total hot water tank volume (L) and water heating power (kW).

3.2 Farm-level Prediction Accuracy

Table 3 Aggregated farm-level prediction results

Target	RPE	CCC	MPE	n
Milk Harvesting	16%	0.96	-3.6%	53
Milk Cooling	12%	0.98	0.5%	53
Water Heating	9%	0.99	0.2%	50

When monthly ANN model predictions were aggregated at the farm-level, milk harvesting RPE equaled 16%, milk cooling RPE equaled 12%, and water heating RPE values equaled 9% (Table 3). Thus, farm-level RPE values reduced by 6%, 12% and 22% points, respectively, when compared to the monthly predictions (section 3.1). CCC values for milk harvesting (0.96), milk cooling (0.98) and water heating (0.99) were all greater than those calculated at the monthly resolution. This suggests a significant improvement between observed and predicted values when predicting at the farm-level. In addition, MPE values of 0.5% for milk cooling, 0.2% for water heating, and -3.6% for milk harvesting further suggest the potential for reduced prediction errors when predicting over a large number of farms.

3.3 Catchment-level Prediction Accuracy

Table 4 Aggregated catchment-level prediction results

Target	Actual (kWh)	Predicted (kWh)	Error
Milk Harvesting	599,776	594,717	0.8%
Milk Cooling	975,873	958,583	1.8%
Water Heating	561,862	551,310	1.9%

A significant reduction in prediction error was observed when monthly ANN model prediction values were aggregated for the entire catchment of study farms. Catchment-level milk harvesting related electricity consumption was estimated with an error of 0.8% (53 study farms), milk cooling electricity consumption with an error of 1.8% (n = 53) and water heating with an error of 1.9% (n = 50), as shown in Table 4.

4. DISCUSSION

When predicting at the monthly resolution all three ANN models resulted in RPE values greater than 20% suggesting poor prediction capability [13]. However, the milk cooling and water heating ANN models resulted in CCC values greater than 0.90 suggesting excellent strength of agreement, while the ANN model predicting milk harvesting electricity use had a CCC value of 0.89 suggesting a substantial strength of agreement [14]. This suggested that the ANN models are responsive to fluctuations in monthly electricity use but contain some absolute prediction error. When predictions were aggregated at farm-level, clear improvements in RPE and CCC values (compared to monthly predictions) were evident, as shown in Table 2 suggesting improved predictions capabilities. This reduced error may be due to a balancing effect between monthly underprediction and overprediction values throughout the year, coupled with the reduced number of datapoints used to calculate farm-level accuracy metrics (50 – 53 study farms).

Predicting water heating electricity use resulted in the largest prediction error (RPE) when predicting at a monthly resolution. This may have been due to water heating being less correlated with the milk production process compared to milk cooling and milk harvesting as water heating was carried out irrespective of milk production. As such, water heating electricity use had to be predicted on months where no milking took place, whereas milk cooling and milk harvesting electricity use was assumed to equal zero and these months were excluded. It is also likely that water heating was carried out more frequently that the frequency required for hot washing (as per input variable 10, Table 1), as hot water was required for other miscellaneous use throughout the farm. However, predicting water heating electricity use at the farm-level resulted in the smallest RPE, again suggesting a balancing effect between over- and underprediction values

5. CONCLUSIONS

Some evidence suggested an improvement in prediction accuracy when predicting at the farm-level compared to the monthly prediction resolution. However, a considerable improvement in prediction accuracy was calculated when predicting at the catchment-level with errors of 0.8%, 1.8% and 1.9% for milk harvesting, milk cooling and water heating, respectively. These results demonstrate the potential effectiveness of the ANN models as macro-level simulation tools for dairy farm electricity consumption.

ACKNOWLEDGEMENT

This work has been funded by the Sustainable Energy Authority of Ireland under the SEAI Research, Development and Demonstration Funding Program 2019, Grant number 19/RDD/453.

REFERENCE

[1] Gerber PJ, Steinfeld H, Henderson B, Mottet A, Opio C, Dijkman J, et al. Tackling climate change through livestock - A global assessment of emissions and mitigation opportunities. Rome: 2013.

[2] FAO. Agricultural Outlook 2018-2027. 2018.

[3] Shine P, Upton J, Sefeedpari P, Murphy MD. Energy consumption on dairy farms : A review of monitoring , prediction modelling and analyses. Energies 2020;13:1288. https://doi.org/10.3390/en13051288.

[4] Shine P, Murphy MD. Over 20 Years of Machine Learning Applications on Dairy Farms : A Comprehensive Mapping Study. Sensors 2022;22. https://doi.org/10.3390/s22010052.

[5] Shine P, Scully T, Upton J, Murphy MD. Multiple linear regression modelling of on-farm direct water and electricity consumption on pasture based dairy farms. Comput Electron Agric 2018;148:337–3. https://doi.org/10.1016/j.compag.2018.02.020.

[6] Todde G, Murgia L, Caria M, Pazzona A. Dairy Energy Prediction (DEP) model: A tool for predicting energy use and related emissions and costs in dairy farms. Comput Electron Agric 2017;135:216–21. https://doi.org/10.1016/j.compag.2017.02.014.

[7] Sefeedpari P, Rafiee S, Akram A, Komleh SHP. Modeling output energy based on fossil fuels and electricity energy consumption on dairy farms of Iran: Application of adaptive neural-fuzzy inference system technique. Comput Electron Agric 2014;109:80–5. https://doi.org/10.1016/j.compag.2014.09.010.

[8] Mhundwa R, Simon M, Tangwe SL. Modelling of an on-farm direct expansion bulk milk cooler to establish baseline energy consumption without milk pre-cooling : A case of Fort Hare Dairy Trust , South Africa. African J Sci Technol Innov Dev 2017;1338:62–8. https://doi.org/10.1080/20421338.2017.1385132.

[9] Edens WC, Pordesimo LO, Wilhelm LR, Burns RT. Energy use analysis of major milking components at a dairy experiment station. Appl Eng Agric 2003;19:711–6. https://doi.org/10.13031/2013.15659.

[10] Shine P, Upton J, Scully T. Machine-learning algorithms for predicting on-farm direct water and electricity consumption on pasture based dairy farms. Comput Electron Agric 2018;150:74–87. https://doi.org/10.1016/j.compag.2018.03.023.

[11] Sefeedpari P, Rafiee S, Akram A. Application of artificial neural network to model the energy output of dairy farms in Iran. Int J Energy Technol Policy 2013;9:82. https://doi.org/10.1504/IJETP.2013.055819.

[12] MacKay DJC. Bayesian Interpolation. Neural Comput 1992;4:415–47. https://doi.org/10.1162/neco.1992.4.3.415.

[13] Fuentes-Pila J, DeLorenzo MA, Beede DK, Staples CR, Holter JB. Evaluation of equations based on animal factors to predict intake of lactating Holstein cows. J Dairy Sci 1996;79:1562–71. https://doi.org/10.3168/jds.S0022-0302(96)76518-9.

[14] McBride G, Bland JM, Altman DG, Lin LI. A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. NIWA Client Rep 2005;45:307–10. https://doi.org/10.2307/2532051.