ARTICLE IN PRESS

Computers and Electronics in Agriculture xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Multiple linear regression modelling of on-farm direct water and electricity consumption on pasture based dairy farms

P. Shine[a], T. Scully[b], J. Upton[c], M.D. Murphy[a,*]

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

A B S T R A C T

An analysis into the impact of milk production, stock numbers, infrastructural equipment, managerial procedures and environmental conditions on dairy farm electricity and water consumption using multiple linear regression (MLR) modelling was carried out. Electricity and water consumption data were attained through the utilisation of a remote monitoring system installed on a study sample of 58 pasture-based, Irish commercial dairy farms between 2014 and 2016. In total, 15 and 20 dairy farm variables were analysed on their ability to predict monthly electricity and water consumption, respectively. The subsets of variables that had the greatest prediction accuracy on unseen electricity and water consumption data were selected by applying a univariate variable selection technique, all subsets regression and 10-fold cross validation. Overall, electricity consumption was more accurately predicted than water consumption with relative prediction error values of 26% and 49% for electricity and water, respectively. Milk production and the total number of dairy cows had the largest impact on electricity consumption while milk production, automatic parlour washing and whether winter building troughs were reported to be leaking had the largest impact on water consumption. A standardised regression analysis found that utilising ground water for pre-cooling milk increased electricity consumption by 0.11 standard deviations, while increasing water consumption by 0.06 standard deviations when recycled in an open loop system. Milk production had a large influence on model overprediction with large negative correlations of −0.90 and −0.82 between milk production and mean percentage error for electricity and water prediction, respectively. This suggested that overprediction was inflated when milk production was low and vice versa. Governing bodies, farmers and/or policy makers may use the developed MLR models to calculate the impact of Irish dairy farming on natural resources or as decision support tools to calculate potential impacts of on-farm mitigation practises.

1. Introduction

GDP (Gross Domestic Product) growth within developing countries is fuelling a forecasted 20% increase in global consumption of milk and dairy products by 2050 (Bruinsma and Alexandratos, 2012). In preparation for the abolishment of the European Union milk quota system in April 2015, the Irish government identified the potential for a 50% increase in milk production by 2020 over 2007–09 levels (DAFM, 2010). With dairy products and ingredients valued at €3bn to the Irish economy in 2014, the increased production should be sensitive to the use of natural resources to ensure the sustainable growth of Ireland's dairy industry (DAFM, 2016). The related impact is twofold: (1) much of Ireland's dairy farm water is supplied by groundwater boreholes to safeguard a consistent, reliable supply of adequate pressure (O'Connor and Kean, 2014). Since 7.42 L of water per litre of milk are consumed on average (Shine et al., 2018), the water demand will rise dramatically in line with milk production which may cause local water shortages during periods of little rainfall, thus placing additional pressure on the public water supply. (2) in Ireland, a strong positive correlation exists between milk production and electricity consumption with 38.84 watt-hours (Wh) per litre of milk consumed on average (Shine et al., 2018). Similarly, the electricity consumption of three Finnish dairy farms varied between 37 and 62 Wh kg^{-1} milk, with milk cooling and milk harvesting being the two largest energy consuming processes (Rajaniemi et al., 2017). Without an effective mitigation strategy, dairy farm electricity costs per litre of milk may increase, as dairy farm infrastructure may not be optimally configured for the increased milk production levels. Similarly, increased electricity consumption during daytime or peak hours may have negative effects on national grid loads as well as on dairy farm electricity costs in a dynamic pricing

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

https://doi.org/10.1016/j.compag.2018.02.020
Received 7 December 2017; Received in revised form 10 February 2018; Accepted 17 February 2018
0168-1699/ © 2018 Elsevier B.V. All rights reserved.

Please cite this article as: Shine, P., Computers and Electronics in Agriculture (2018), https://doi.org/10.1016/j.compag.2018.02.020

environment (Upton et al., 2013).

Milk price volatility and environmental constraints are forcing farmers to produce milk at lower costs with a lower overall environmental footprint. Empirically predicting overall electricity and total dairy farm actual water consumption (leak inclusive) may allow for an efficient and comprehensive means of calculating dairy farm electricity and water (E&W) consumptions as metering equipment and/or large input mathematical models are replaced by a small number of high prediction yielding, empirically derived coefficients, allowing consumption values to be easily calculated. The ability to efficiently predict dairy farm E&W consumption supports the sustainable growth of Ireland's dairy farms, which is of benefit to governing bodies, policy makers, dairy farmers and Irish dairy industry stakeholders.

Current state-of-art electricity prediction on Irish dairy farms is mechanistic in nature (Upton et al., 2014), capable of predicting total CO_2 emissions, electricity costs and consumption, predicting the latter to < 10%. Large scale mechanistic modelling would require an exhaustive data collection process to collect specific data related to plate heat exchanger milk: water ratios, water temperatures, milking times, installed lighting capacity etc. (Upton et al., 2014). On the other hand, an empirical model for dairy farm electricity consumption would replace this large number of input variables with a small number of highly predictive numerical coefficients. This approach was undertaken in Southern Italy, where a MLR model was developed for predicting annual dairy farm related electricity consumption to within 11.4% utilising 285 dairy farms (Todde et al., 2017). Similarly, MLR models were developed to predict the individual and combined electricity consumption of the main electrical components (vacuum pumps, refrigeration compressors, water heaters, and air compressor) of a single farm using 14 years of metered data (Edens et al., 2003). Edens et al. (2003) were capable of explaining 62% of the variability of the combined electricity consumption of the main electrical components. Concurrently, Edens et al. (2003) concluded that quantity as opposed to the quality of milk produced had a far greater impact on electricity consumption with the number of cows milked and ambient temperatures having lesser statistical impacts.

Research regarding dairy farm water prediction has primarily focused on developing linear regression models for daily cow drinking water requirements within confined dairy systems (coefficient of determinations ranging from 0.39 to 0.75) (Cardot et al., 2008; Meyer et al., 2004; Murphy et al., 1983). However, linear regression functions have been developed for predicting annual Irish dairy farm green and blue water volumes to 11.3% and 3.4%, respectively (Murphy et al., 2017). These studies required meticulous data collection including concentrates fed, direct water demand (for blue water prediction), dry matter (DM) feed intake, live weight and sodium intake. In contrast to a meticulous data collection approach, Higham et al. (2016) utilised variables related to milk production, stock, environmental conditions and infrastructural equipment from 23 New Zealand pasture based dairy farms. Using this approach, they developed a partial least square regression model (r = 0.90) for predicting total daily water consumption (drinking plus parlour).

MLR modelling has also been applied within the agricultural domain for forecasting milk production with accuracies ranging from 9% to 26% (Murphy et al., 2014; Sharma et al., 2006; Zhang et al., 2016). MLR models utilising data related to milk production, stock, farm infrastructure, managerial processes and environmental conditions for dairy farm E&W prediction can provide; (1) a decision support tool for dairy farmers and/or policy makers to calculate the impact of potential on-farm mitigation practises. (2) governing bodies with the ability to conduct macro level E&W analysis or water risk assessments as variables may be gathered on a large scale without the requirement of specialised equipment. (3) governing bodies and state agencies such as the Department of Agriculture, Food and the Marine (DAFM) and Bord Bia with the means of calculating the impact of Irish dairy farming on natural resources for sustainability reporting and marketing Irish dairy products abroad (DAFM, 2016).

This work utilised E&W consumption data collected from 58 Irish commercial dairy farms and corresponding data related to milk production, stock, farm infrastructure, managerial processes and environmental conditions. Utilising this data, the objectives of this work were to: (1) develop MLR models for both electricity and water (leak inclusive) consumption from an initial selection of variables, by employing a range of data mining techniques to extract variable subsets which offer the greatest prediction accuracy on unseen consumption data. (2) analyse the impact of the optimum variables on E&W consumption through a standardised regression analysis. (3) analyse the monthly prediction bias of each MLR model to determine factors, which may influence model performance.

2. Materials & methods

In this study, data were acquired via both automated and manual recording of E&W consumption through Teagasc, Moorepark (Cork, Ireland). In total, 58 Irish pasture based dairy farms were monitored throughout the period 1st Jan 2014 – 31st May 2016. Electricity consumption was monitored on 56 farms (autonomous = 55, manual = 1) while total water consumption was monitored on 51 farms (autonomous = 20, manual = 31). Farms that had part of total consumption acquired through manual reporting were classified as manually recorded dairy farms. The dairy farms used in this study had a mean herd size of 116 cows and annual milk production of 621,702 litres in 2015 as described by Shine et al. (2018). Data utilised for MLR model development are summarised in Table 1 below. This data includes milk production, cow numbers, E&W consumption data and related key performance indicators, at a monthly resolution. On average, 51,421 L of milk was produced per month, which resulted in a consumption of 2094 kWh of electricity and 361 m^3 of water on average. Regarding dairy farm infrastructure, 47 (84%) of the dairy farms utilised for developing the electricity MLR model employed a direct expansion bulk

Table 1
Population descriptions for monthly electricity and water consumption, milk production, dairy cows and related key performance indicators.

Variable	Unit	Min	Mean	Median	Max	IQR	SD	SEM
Milk yield	Litre	213	51,421	48,016	204,756	42,407	32,452	863
Dairy cows	n	28	114	102	300	50	41	1
Electricity	kWh	199	2,094	1,818	7,786	1,350	1,094	31
	Wh L_m^{-1}	8.10	73.19	39.82	3,314.64	31.68	164.97	4.70
	kWh Cow^{-1}	1.98	18.35	18.14	45.36	7.78	6.02	0.17
Water	m^3	51	361	308	1575	217	218	7
	L_w L_m^{-1}	1.67	13.42	6.63	542.22	6.82	34.18	1.14
	m^3 Cow^{-1}	0.51	3.29	2.95	23.50	1.83	1.90	0.06

IQR = Inter-quartile range, SD = Standard deviation, SEM = Standard error of the mean.
Wh L_m^{-1} = Watt-hours per litre of milk. kWh Cow^{-1} = Kilowatt-hours per dairy cow.
L_w L_m^{-1} = Litres of water per litre of milk. m^3 Cow^{-1} = Cubic meter of water per dairy cow.

tank with the remaining nine (16%) employing an ice bank milk cooling system. Concurrently, 46 of the 51 farms (90%) utilised for the development of the water MLR model pre-cooled milk through a plate cooler, either through ground water or ice-cold water (through employing an ice chiller). Of these 46 dairy farms, plate cooler water was recycled throughout the farmyard in an open loop system on 40 farms while a further three farms recycled plate cooler water (either ground or ice cold water) through a closed loop. A further three farms did not recycle plate cooler water.

Data were considered to have noise present due to the inherent nature of data recording where there are many exogenous factors affecting consumption readings. The water consumption data in particular were susceptible to noise due to: (1) over 60% of farms manually reporting consumption on a monthly basis, which prevents adjustment in the event of a leak and (2) some exogenous factors that affect cow drinking water on pasture based dairy farms were outside the scope of this research. Factors such as live weight and sodium intake affecting cow drinking water as well as soil type, altitude and stocking rates affecting the dry matter (DM) intake of a cow's diet and thus cow drinking water (Hanrahan et al., 2017) were not considered.

The methodology flow from the initial variables considered (Table 2) to the performance calculation of each variable subset allowing the optimum variable subset to be selected is outlined in Fig. 1. The process may be categorised into three sub-sections: (1) a variable selection section where initial variables considered are presented with the methodology employed to extract those, which yield high predictive power (Section 2.1). (2) a section discussing All Subsets Regression (ASR) where MLR models were developed for all possible subsets of those selected variables (Section 2.2). (3) a section describing methods utilised for MLR validation and performance accuracy calculations (Section 2.3). All statistical processes were carried out using MATLAB 2016b.

2.1. Variable selection

In total, 15 potential variables were analysed for electricity prediction modelling while 20 variables were analysed for water prediction modelling as presented in Table 2. These variables were related to; milk production and stock (variable IDs: 1–3), infrastructural equipment (4, 5, 9:21 and 24:27), managerial procedures (6 and 23) and environmental data (7, 8 and 22). Milk production data throughout the analysis period was recorded (litres) by the milk processors. Herd size (total on-farm dry and lactating dairy cows) data were attained for each farm through the Irish Cattle Breeding Federation (ICBF, 2016). Monthly mean minimum temperature (°C), mean maximum temperature (°C) and precipitation levels (mm) were obtained from Met Éireann (Met Éireann, 2017) for five metering stations with data from the nearest neighbouring station used for each farm. A Teagasc technician completed a survey on each farm to acquire data related to on-farm infrastructural equipment and managerial processes. The frequency of hot washing milk lines (HzHW) was normalised to the number of hot washes per month based upon a 30 day month.

ASR was employed to determine the subset of variables, which best predict E&W consumption, as described by Elliot et al. (2016). ASR calculated the prediction accuracy of all possible subsets of variables, and selected the subset which resulted in the greatest accuracy through *k*-fold cross validation (discussed in Section 2.3) based upon a pre-determined criterion. However, it was too computationally expensive to analyse all possible subsets of 15 and 20 variables for E&W, respectively. Thus, all initial variables underwent a univariate selection process before model development to reduce the variable space by excluding variables which add little predictive power. It was then feasible to carry out ASR to extract the most accurate variable subset.

2.1.1. Univariate variable selection

A univariate variable selection approach was undertaken to identify highly correlated variables for linear prediction. For categorical variables (such as whether an ice bank or direct expansion milk cooling tank was used), a difference in monthly consumption per dairy cow (minimum effect size = 0.1) and maximum P value of 0.001 was required between categories for selection. The effect size measured strength of the standardised difference between population means, expressing the difference in units of standard deviations (Daly and Cohen, 1988). Concurrently, a test for a relationship between monthly E&W consumption and continuous variables such as milk production was carried out with a maximum p value limit of 0.001 employed alongside a minimum correlation coefficient (r) of 0.4 required. This method extracts highly correlated variables for MLR development as established and discussed in previous research (Shine et al., 2018).

Multicollinearity (Elliot et al., 2016) is a correlation existing between the predictive variables whereby models developed from collinear data may lead to erroneous system analysis (Akhil and Kang, 2013). To analyse the presence of multicollinearity within the set of variables, collinear variables were removed through stepwise iterative Variance Inflation Factor (VIF) (Eq. (1)) analysis as in previous literature in this research space (Todde et al., 2017).

$$VIF_j = \frac{1}{1-R_j^2} \tag{1}$$

where the VIF_j is the VIF value for the jth variable. VIF was calculated from the reciprocal of the inverse of R^2 from the regression equation developed from all other variables. The VIF for all included variables was calculated and variables with the largest VIF above the threshold of

Table 2
List of initial dairy farm variables for model development.

Variable ID	Variable description	No. of farms		Variable ID	Variable description	No. of farms	
		Electricity	Water			Electricity	Water
1	Number of dairy cows	56	51	15	Total number of scrapers	56	–
2	Number of lactating cows	56	51	16	Supply water flow rate	–	44
3	Milk production	56	51	17	Automatic parlour washing	–	51
4	Number of parlour units	56	51	18	Rain water collection	–	51
5	Total water heater volume	54	50	19	Plate cooler water for recycled in open loop	–	51
6	Frequency of hot Washing	56	51	20	Winter building troughs leaking	–	51
7	Mean minimum ambient temperature	56	51	21	Field troughs ballcock high or low flow	–	49
8	Mean maximum ambient temperature	56	51	22	Precipitation	–	51
9	Ice cold water utilised for pre-cooling	56	–	23	Time spent parlour washing daily	–	42
10	Ice bank or direct expansion bulk tank	56	–	24	Field trough pipe diameter	–	46
11	Ground water utilised for pre-cooling	56	–	25	Wash pump hose type	–	51
12	Total bulk tank volume	56	–	26	Field troughs average volume	–	49
13	Number of Air Compressors	56	–	27	Wash down tank capacity	–	43
14	Number of vacuum pumps	56	–				

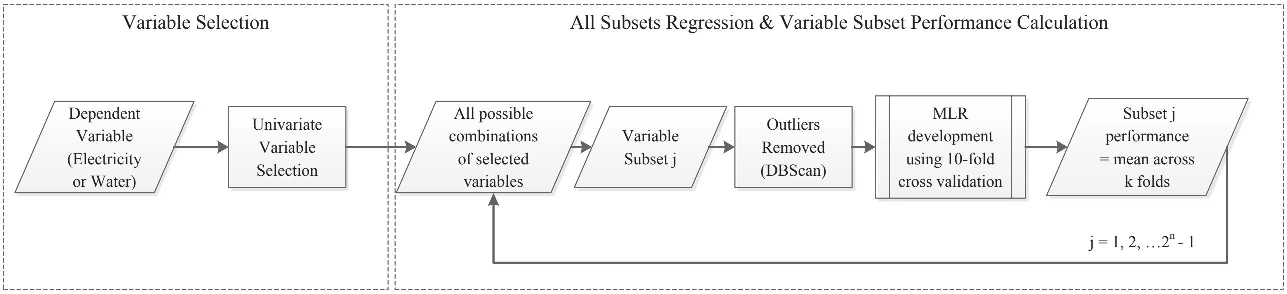

Fig. 1. Schematic of methodology flow from raw data to model validation and performance calculation.

5 removed (Herrig et al., 2015). This process was repeated until the VIF value for all variables were below the threshold of five.

2.2. All subset regression

Once the univariate variable selection process was performed, ASR was carried out to allow for the selection of optimum variable subsets. ASR is an exhaustive approach, creating separate MLR models for all possible subsets of variables without repeats. The number of possible variable subsets to be analysed (Eq. (2)) doubles with every additional variable considered.

$$C = 2^n - 1 \tag{2}$$

where, C represents the total number of possible subsets of n predictor variables.

2.2.1. Data pre-processing

No pre-processing methods were employed to replace missing data points. Hence, all missing data iterations (dependent or predictor variables) were removed, resulting in a clean dataset for variable selection. Concurrently, non-milking months (data points with zero milk production) were excluded in the model development to reduce the coefficient of variation of milk production and thus inherent noise within the E&W datasets. Prior to model development, data from the jth subset (illustrated in Fig. 1) were pre-processed to remove erroneous data points (outliers), which may have had a negative effect on coefficient calculations and thus on model accuracy. The presence of outliers can be attributed to a range of factors such as meter faults, leakage, human error, etc. For each subset of variables, data related to each variable was standardised to a mean value of zero and standard deviation of one before the dataset was checked for outliers using the Density Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm (Ester et al., 2010).

DBSCAN marks three types of points: core (z), border (y) and outliers (x) in a multidimensional space (Fig. 2 presents these concepts in two-dimensional space). A core data point has at least a specified number (*MinPts*) of neighbouring points within a fixed radius ε. A border data point has less than *MinPts* neighbouring data points but lies within ε radius of a core data point. Outliers are neither core nor border points, which lie alone in low-density regions in the geometrical space. The DBSCAN algorithm selects a core data point and begins to form a cluster by identifying all other core and border data points that are reachable from it.

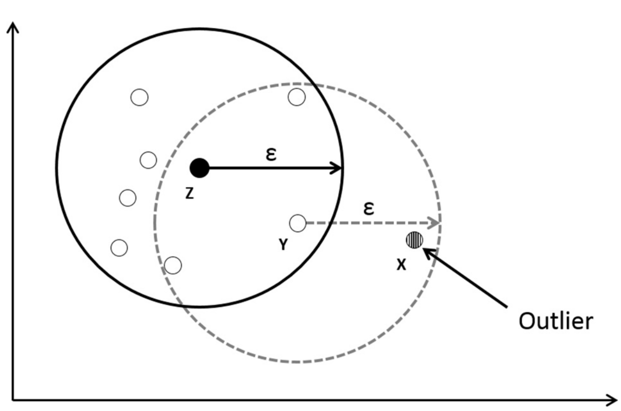

Fig. 2. Determining outliers using DBSCAN in a 2-dimensional space.

The selection of appropriate values for ε and *MinPts* parameter values is an important consideration when using DBSCAN. Increasing the value of ε typically leads to an increase in the number of data points in each cluster. On the other hand, the greater the value of *MinPts* required to form a cluster, the greater the proportion of data identified as outliers. The desired ε parameter value was calculated using a k-dist method (Ester et al., 2010). Firstly, a *MinPts* parameter value of four was prescribed for all subsets, as k-dist graphs for k values above four have been found to not significantly differ and require considerably greater computational power (Ester et al., 2010; Malathi and Rajarajeswari, 2014). The k-dist method calculated the distance from each data point to its *MinPts* nearest neighbour. Distances were sorted (smallest to largest), curve plotted and a line calculated from maximum to minimum distance points. The knee point (where a valley occurs along the k-dist curve) corresponded to the desired ε value. The knee point was defined as the point with the maximum perpendicular distance between the line and the graph (Satopää et al., 2011).

2.2.2. MLR development

MLR models describe the linear relationship between multiple predictor variables for the prediction of a single dependent variable, as is in Eq. (3). MLR equations were developed for each month and described the best fit line which minimised the sum of the squared error of the vertical deviations from each observed data point to the line.

$$Y_i = \varepsilon_i + \alpha_{1i}X_{1i} + \alpha_{2i}X_{2i} + \ldots + \alpha_{ni}X_{ni} \tag{3}$$

where Y_i is electricity or water consumption for month i, ε_i is the corresponding intercept term, $a_{1i}, a_{2i}, \ldots, a_{ni}$ are the prediction coefficients and $X_{1i}, X_{2i}, \ldots, X_{ni}$ are the predictor variable values for the i^{th} month for the n^{th} variable. Linear coefficients were calculated on a monthly basis coinciding with the availability of monthly milk production and stock data.

An important criterion for MLR model coefficients is the normality of standardised model residuals about zero (Ngo, 2012). Non-normal residuals are common when coefficients are derived from non-normal data. Similarly, a small number of data points may push or pull the best fit line in a particular direction, skewing prediction values. Thus, if residuals failed a one-sample Kolmogorov-Smirnov test against a normal distribution ($p < 0.001$), values outside three standard deviations of the mean were excluded, coefficients re-calculated and residuals re-analysed. If residuals remained non-normal, the standard deviation limit for excluding outliers decreased in 0.25 increments thereafter until residual normality was found. This ensured the normality of standardised model residuals about zero.

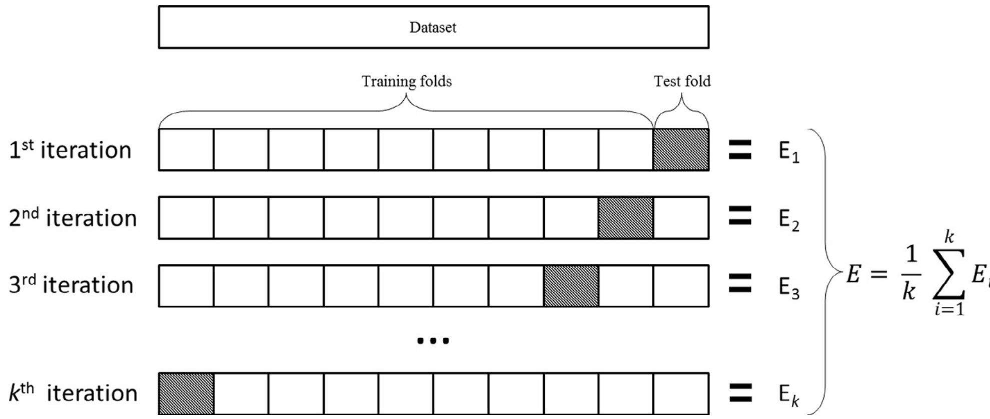

Fig. 3. Overview of *k*-fold cross validation method.

2.3. MLR validation and performance

The prediction accuracy of each variable subset was calculated through a stratified *k*-fold cross validation ($k = 10$) approach, as depicted in Fig. 3. Data were split into *k* sections or 'folds' of approximately equal size, where each fold contained an equal number of readings for each month. Cross fold validation is an iterative processes whereby the model performance is evaluated for each fold. For the *i*th iteration ($i = 1$: k), MLR coefficients were calculated on a training dataset consisting of data from *k-1* folds (full dataset excluding the i^{th} fold) while prediction accuracy was assessed on a test set consisting of the remaining fold (test set of unseen data (i^{th} fold)). This process was repeated k times, until the prediction accuracy was calculated on each of the *k* folds. Overall model prediction performance criteria (*E*) was calculated as the mean value across all folds. Conducting *k*-fold cross validation for each variable subset allowed for the variable subset, which offered the greatest accuracy in predicting unseen electricity consumption to be selected.

Model bias was evaluated according to mean percentage error (MPE (%)) and mean square prediction error (MSPE) (decomposed into mean bias (MB), line bias (LB) and unexplained random variation (RV) percentage according to their proportion contribution to MSPE) (Bibby and Toutenburg, 1977). Model precision was evaluated according to root mean squared error (RMSE) (square root of MSPE) (Bibby and Toutenburg, 1977), relative prediction error (RPE (%)) (RMSE expressed as a percentage of the mean actual electricity or water value) (Rook and Gill, 1990) and concordance correlation coefficient (CCC) (Lin, 2016). These criteria were employed in cognate studies analysing model prediction in agricultural applications (Baudracco et al., 2013; Hanrahan et al., 2017; Todde et al., 2017; Upton et al., 2014). RPE computes the ratio of the absolute error with respect the mean actual value while CCC measures the strength of agreement between actual and predicted values. Models with RPE values greater than 20% are considered to have poor prediction capability, values between 10% and 20% suggest acceptable prediction and values lower than 10% indicate satisfactory prediction performance (Fuentes-Pila et al., 1996). CCC values greater than 0.90 suggest the strength of agreement between actual and predicted values is excellent, values between 0.80 and 0.90 considered substantial, while values between 0.65 and 0.80 are considered moderate and poor if lower than 0.65 (McBride et al., 2005). Model selection was based upon the subset of variables which offered the greatest accuracy with respect to RPE.

The results section may be divided into three sub-sections: (1) descriptive statistics for the most accurate variable subsets for E&W consumption. (2) a standardised regression analysis to determine the effect of each input variable on E&W consumption prediction. (3) MPE is presented according to each month with respect to mean and standard deviation of milk production. MPE values for each month were normalised according to the mean monthly E&W consumption values as large percentage error values during months of low consumption may have low impacts regarding annualised consumption prediction errors.

3. Results and discussion

3.1. MLR models

3.1.1. Electricity

The univariate variable selection method reduced the variable space from 15 to ten highly correlated variables. Although highly correlated with electricity consumption, the VIF analysis identified the number of milking cows to be collinear (VIF > 5) with the remaining variables and was thus excluded from further analysis. Although both the utilisation of ice-cold water for milk pre-cooling and the number of parlour units were found to impact electricity consumption, these variables were not included in the variable subset which optimised the prediction of unseen electricity consumption. The most accurate MLR model configuration for electricity consumption prediction contained seven variables, as shown in Table 3a. This subset comprised of the total number of dairy cows, milk production, the total water heater volume, HzHW, whether an ice bank (IB) or direct expansion (DX) bulk tank was employed, whether ground water (gW) was utilised for pre-cooling and the number or air compressors. This combination of variables shows similarities to research carried out by Edens et al. (2003) as both models found milk production to be a key input variable. However, the number of cows milked and the monthly max temperature were input variables in Edens et al. (2003) and were excluded in the optimum MLR

Table 3
Electricity and Water MLR prediction accuracy.

Variable IDs	RPE	CCC	MPE	r	RMSE	MSPE	MB	LB	RV	n	Outliers removed
					(a) Electricity						
1,3,5,6,10,11,13	26.1%	0.84	−8.4%	0.85	543 kWh	295,605	1%	0%	99%	1155	2.3%
					(b) Water						
1,3,4,17,19,20	49.4%	0.47	−18.1%	0.52	175 m^3	30,839	1%	0%	99%	892	0.9%

in this study. The optimum MLR model in this study resulted in a RMSE value of 543 kWh, capable of predicting monthly electricity consumption to within 26.1% (RPE) with an CCC value of 0.85 and MPE value of -8.4%, as shown in Table 3a. Consequently, this model was above the 20% accuracy threshold to be considered having acceptable prediction capability with regard its RPE. However, the strength of agreement (CCC) between actual and predicted values was considered substantial. Good model correlation with poor absolute error suggests that: (1) prediction may be overly sensitive to certain variable coefficients causing over and/or under shooting and (2) steady state error effects (consistent under/overprediction with good response) may be the root cause of the poor absolute prediction capabilities of the MLR model. This may be explained by the coarse nature of some input variables such as whether ground water is used for pre-cooling (yes|no) as opposed to an exact water to milk ratio. In comparison to cognate studies, the electricity model (RPE = 26.1%) was found to be far less accurate than that developed by Todde et al. (2017) (RPE = 11.4%), however, this discrepancy may be explained by the different prediction resolutions, validation methodology and incorporation of a cubed term. Concurrently, the electricity model (r = 0.84) was found to be more accurate than that developed by Edens et al. (2003) (r = 0.79), even with this study calculating accuracies on unseen data.

The calculated MLR coefficients for the developed electricity model can be found in Table 4 below.

3.1.2. Water

The univariate variable selection method reduced the variable space from 20 to seven highly correlated variables. No variable was excluded due to multicollinearity. However, although the number of milking cows was highly correlated with water consumption, this variable was not included in the variable subset, which optimised the prediction of unseen water consumption. The addition of the number of milking cows, to the optimum subset of variables may have resulted in an overfitting model, thus reducing its predictive accuracy on unseen data. The most accurate MLR model configuration for water consumption prediction contained six variables: the total number of dairy cows, milk production, the number of parlour units, automatic parlour washing, whether gW was utilised for pre-cooling in an open loop system and whether water troughs were reported to contain leaks, as shown in Table 3b. The main difference between the water model developed in this study and that of Higham et al. (2016) is the lack of any meteorological variable inputs in the model developed in this study. The model developed by Higham et al. (2016) included: maximum daily temperature, evapotranspiration, solar radiation, milk solids, milk volume and the milking parlour type. Although mean minimum and maximum ambient temperature and precipitation were considered, no meteorological variable resulted in a univariate correlation above the required threshold to be considered for further analysis. Concurrently, both models found milk production to be a key input variable. The optimum MLR water model in this study resulted in a RMSE value of 175 m^3, capable of predicting monthly water consumption to within 49.4% (RPE) with an CCC value of 0.52 and MPE value of −18.1%. Consequently, this MLR model exhibits poor predictive capability with regard to the RPE (RPE > 20%) and CCC (CCC < 0.65) precision scales. In comparison to Higham et al. this water model predicted total water consumption less accurately; however, this may be due to differences in resolution and/or Higham et al. adjusting for water leakage as a pre-processing step.

The calculated MLR coefficients for predicting water consumption on dairy farms can be found in Table 5 below.

3.2. Regression analysis

Standard regression coefficients (Fig. 4) were calculated by transforming each variable vector to a mean value of zero and standard deviation of one, thus eliminating units. The standardised regression coefficients were calculated for the entire dataset to analyse the averaged impact of each variable throughout the year as opposed to the impact on a monthly basis. Standardised coefficients measure how many standard deviations E&W will change per standard deviation change of the predictor variable. This allows for comparisons to be made on the impact of each variable on E&W consumption (Schroeder et al., 1996). A variable with a positive standard regression coefficient represents a subsequent increase in consumption with an increase in the value of the variable input and vice versa. Note: for categorical variables, the binary input of each category (0 or 1) impacts the sign of the regression coefficient.

3.2.1. Electricity

As shown in Fig. 4a, the total number of dairy cows has the largest effect on electricity consumption with one standard deviation change resulting in 0.42 standard deviation change in electricity consumption. Milk production had the second largest effect with one standard deviation change resulting in 0.29 standard deviation change in electricity consumption. These results reflect recent findings whereby variances in electricity consumption outside of milking processes are best explained by stock numbers and milk production volumes (Shine et al., 2018). Similarly, milk harvesting and cooling are two of the largest energy consuming processes on dairy farms (Murgia et al., 2013; Rajaniemi et al., 2017). The remaining five variables had the following effect on electricity consumption (in decreasing order of impact): the HzHW (0.15), the number of air compressors (0.13), whether ground water (gW) was utilised (Yes = 0, No = 1) for pre-cooling (0.11), whether an IB or DX milk tank (DX = 0, IB = 1) was employed (0.09) and the water heater volume (0.08). Whether gW was utilised for pre-cooling and whether an IB or DX milk tank was employed for milk cooling were

Table 4
MLR model coefficients for predicting monthly electricity consumption on dairy farms.

Equation input	No. of dairy cows n	Milk yield m^3	Cooling system DX = 0, IB = 1	Milk pre-cooling Yes = 0, No = 1	No. of air compressors n	HzHW Hot washes/month	Total water heater volume L	Constant –
Jan	9.73	26.30	−664.67	−17.21	−79.28	24.82	−0.54	251.36
Feb	13.34	9.66	70.02	110.83	69.72	28.20	−0.61	−69.41
Mar	12.21	9.64	302.18	240.52	273.16	17.48	0.23	5.22
Apr	10.01	8.27	409.51	184.04	316.18	10.52	1.18	−94.57
May	11.05	4.76	421.59	391.02	401.69	8.85	1.93	−187.64
Jun	9.61	7.88	429.57	427.05	366.19	8.16	1.60	−289.57
Jul	13.82	2.27	288.78	467.28	349.60	8.08	1.72	−341.83
Aug	8.58	11.30	187.61	439.87	343.01	13.11	1.14	−340.06
Sep	5.55	15.31	205.52	410.73	292.61	9.11	1.55	−294.28
Oct	6.98	14.30	124.08	335.28	267.98	9.21	1.68	−253.44
Nov	6.06	30.47	30.84	246.73	191.45	13.22	0.38	−192.87
Dec	4.94	27.40	−207.29	233.63	171.06	20.66	1.03	−90.16

MLR coefficients for each input variable, for each month are presented above. Note input variable units below each input heading.

Table 5
MLR model coefficients for predicting monthly water consumption on dairy farms.

Equation input	No. of dairy cows	Milk yield	No. of parlour units	Auto parlour washing	Open loop milk pre-cooling system	Troughs leaking	Constant
	n	m^3	n	No = 0, Yes = 1	Yes = 0, No = 1	No = 0, Yes = 1	–
Jan	−1.86	8.50	34.08	−79.97	58.96	158.67	−77.92
Feb	−1.46	4.96	13.33	125.01	−16.51	183.65	123.33
Mar	−0.34	0.47	13.07	142.20	−38.30	200.30	173.63
Apr	1.96	−0.83	9.29	97.35	−41.96	187.38	71.23
May	1.14	1.11	2.43	64.07	−51.74	220.80	112.08
Jun	−0.05	0.77	11.99	77.30	−68.01	112.93	156.48
Jul	0.26	1.02	13.19	44.74	−48.02	18.96	104.87
Aug	0.59	0.21	7.23	77.64	−69.98	71.61	163.51
Sep	2.73	−2.32	−1.55	57.82	−71.18	39.82	170.33
Oct	−0.16	4.32	−1.09	0.59	−79.88	−4.70	161.24
Nov	−0.56	4.59	2.93	136.13	6.40	149.39	133.16
Dec	−2.03	6.43	11.52	174.57	−42.20	0.41	210.99

MLR coefficients for each input variable, for each month are presented above. Please note input variable units below each input heading.

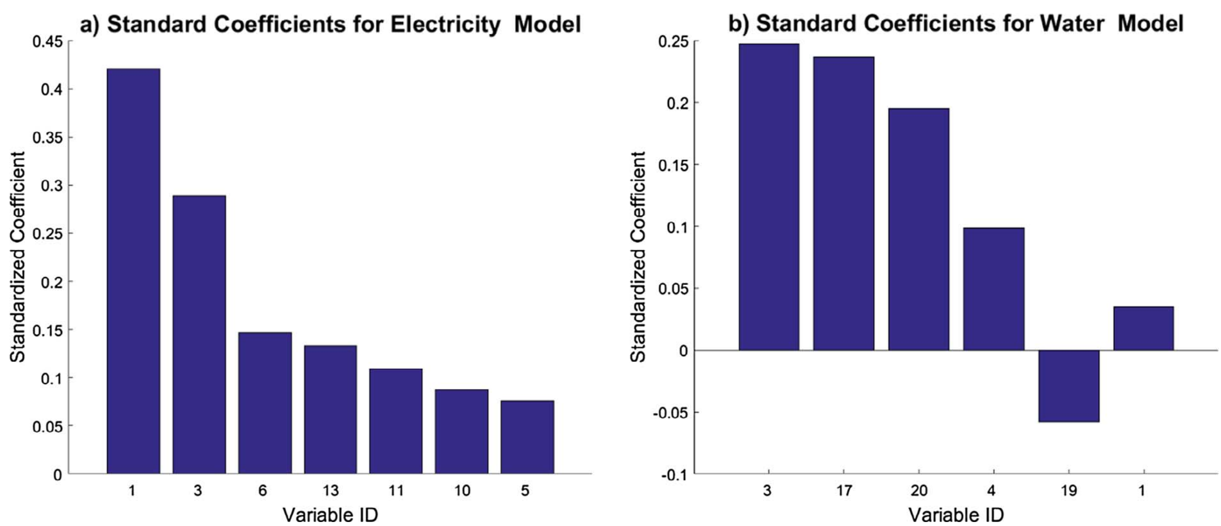

Fig. 4. Variable standardised coefficients for electricity and water models. Explanations of variable IDs are presented in Table 2.

categorical variables and thus required binary inputs. The binary input associated with each category (e.g. DX = 0, IB = 1) determined whether a positive or negative standard coefficient was calculated. However, the absolute impact remained constant. A large correlation coefficient ($r = 0.51$, $p < 0.001$) has previously been reported between dairy farm electricity consumption and HzHW, hence its predictive impact on consumption (Shine et al., 2018). Milk cooling was the highest electricity consuming process responsible for 29% of overall demand (Shine et al., 2018). Thus, data related to milk cooling (whether an IB or DX bulk tank was employed on farm and whether gW is utilised for milk pre-cooling) was seen to impact electricity consumption. These results suggest that IB milk cooling systems consume greater amounts of electricity than DX systems as a positive standard coefficient was calculated for the binary input of one for IB systems. This may also be interpreted as dairy farms with DX milk cooling systems consumed 0.09 standard deviations less electricity than their IB counterparts. This result was reflective of recent findings where Ice bank systems have been found to consume 32% more electricity for milk cooling than standard DX systems due to the high energy demand of its refrigeration unit (Shine et al., 2018). Concurrently, farms that did not pre-cool milk with water consumed a greater magnitude of electricity than farms without pre-cooling systems as a positive standard coefficient was calculated for the binary input of one for farms which did not pre-cool milk. This finding was reflective of recent findings where farms which employ gW for pre-cooling milk can reduce milk cooling electricity consumption by 21% (Shine et al., 2018). Air compressors may be used on farms for a variety of reasons including: powering milking machines, cleaning equipment and piping systems and powering pneumatic tools. However, previous studies have reported the contribution of air compressors to overall consumption to be much lower (< 3%) than the quantified impact on consumption found in this study (Shine et al., 2018; Upton et al., 2013). Thus, the impact of air compressors on electricity in this regression analysis may be inflated. Water heating is responsible for 20% of total electricity consumption on Irish dairy farms (Shine et al., 2018). Electricity consumption attributed to water heating may be influenced by factors such as water tank volume and ambient temperature. In this study, variances in water heating electricity consumption were explained by water heater tank volume as naturally, greater tank volumes require a greater magnitude of electricity to heat to a specific temperature.

3.2.2. Water

As shown in Fig. 4b, milk production had the largest impact on water consumption with one standard deviation change resulting in a 0.25 standard deviation change in water. This may be explained by dairy cows requiring large volumes of drinking water during lactation for milk production (Cardot et al., 2008). As mentioned in Section 3.1.2, milk production was also found to be a key input variable in the MLR model developed by Higham et al. (2016). Automatic parlour washing (No = 0, Yes = 1) had the second largest effect with one standard deviation change resulting in 0.24 standard deviation change in water consumption. Automatic parlour washing systems are labour efficient; however, these results suggest a significant increase in water consumption as a result.

The remaining four variables had the following effect on water consumption (in decreasing order of impact): whether winter building troughs were reported to contain leaks (No = 0, Yes = 1) (0.20), the number of parlour units (0.10), whether gW was utilised for pre-cooling in an open loop system (Yes = 0, No = 1) (−0.06) and the total number of dairy cows (0.03). Winter building troughs which contain leaks will clearly consume greater volumes of water as water is lost on a consistent basis. A moderate correlation between water consumption and the number of parlour units ($r = 0.38$, $p < 0.001$) has previously been reported (Shine et al., 2018). Increasing the number of parlour units will require a greater volume of water for washing of piping systems and for washing an increased parlour floor area. Not utilising gW for milk pre-cooling in an open loop system (recycling via external wash down tank after milking) resulted in decreased water consumption. This is reflective of recent literature reporting that farms which recycled plate cooler water in open loop systems consumed 41% greater volumes of water within the milking parlour than farms which do not pre-cool (Shine et al., 2018). This was due to farms that recycle plate cooler water in open loop systems consuming 119% more wash water then those farms without pre-cooling facilities. Although a moderate correlation existed between dairy cow numbers and water consumption, a correlation also existed between the number of dairy cows and milk production. As milk production had the largest impact on water consumption, it is likely that variances due to the number of dairy cows was contained in the milk production data and the low standard coefficient presented here is due to non-lactating dairy cows consuming water.

3.3. Monthly prediction performance

The overprediction tendencies of the E&W MLR models per month is shown in Fig. 5 with associated mean and standard deviations of milk production across the training datasets. For ease of reading, the right side y-axis (MPE) has been inverted. Prediction tendency is represented by MPE where negative values represent overprediction, the absolute value of which represents the scale of over or under prediction. Both the mean and standard deviation are presented to show the relative disparity of the data for each month. Both values allow for the coefficient of variation (ratio of standard deviation to the mean) to be calculated (Brown, 1998). The greater the standard deviation to mean milk production ratio, the greater disparity in the data.

The electricity model overpredicted consumption by 8.4% (MPE = −8.4%) on average (Table 3a). Overprediction bias was amplified during the winter months, with January (−22%) and February (−17%) the two least accurate months whereas June (−3%) and July (−4%) were the two most accurate months, as shown in Fig. 5. The water model overpredicted consumption by 18.1% on average (Table 3b). Overprediction bias was amplified in January (−50%) and December (−29%) the two least accurate months whereas August (−8%) and July (−10%) were the two most accurate months (Fig. 5).

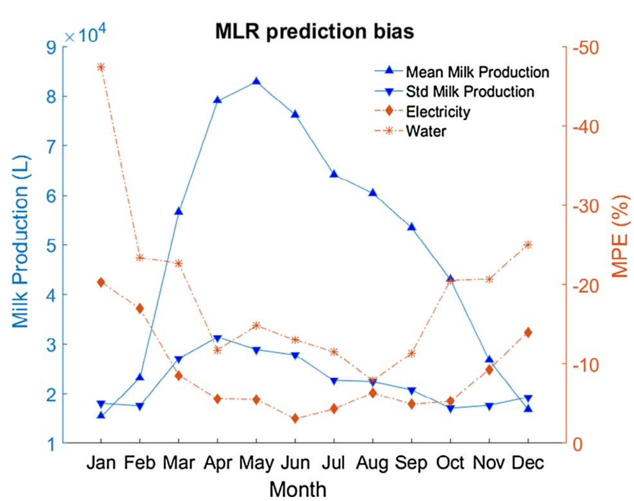

Fig. 5. Electricity and water MLR model bias and accuracy according to month.

Regarding electricity, overprediction bias in months January, February, November and December was amplified to a mean MPE value of −14.5%. However, this value was improved by 62% (62% closer to zero) during the primary milking months (March – October) to a MPE value of −5.6%. Similarly for water, overprediction bias in months January, February, November and December was amplified to a mean MPE value of −25.6%, improving by 44% during the primary milking months to a MPE value of −14.3%. Regarding the prediction of electricity consumption, a Pearson's correlation coefficient value of −0.90 was calculated between milk production and MPE. Similarly, a Pearson's correlation of −0.82 was calculated between milk production and MPE regarding water consumption prediction. These negative correlations suggest that model overprediction was inflated when milk production was low. During the March to October period, the standard deviation of milk production was equal to 39% of mean milk production (coefficient of variation), on average. During this period, most farms are producing milk at close to maximum capacity, thus decreasing the coefficient of variation. In contrast, for the January, February, November and December period, the coefficient of variation was equal to 93% of mean milk production as some farms produced little milk and some still produced large volumes. This may be attributed to a phase shift between farms where some milk all year round, some dry off earlier/later than others and some farms begin milking earlier/later. This resulted in increased electricity and water consumption variance between farms, which in turn lead to an increase in the overprediction percentage error of farms with low milk production.

3.4. Discussion overview

The electricity consumption model was 47% more accurate (RPE = 26%) than water the consumption model (RPE = 49%). Electricity is a more deterministic type of consumption where consumption may only take place when demanded from an electrical load. On the other hand, water demand may be affected by leakage of water pipelines or by exposure to irradiance and wind shear, increasing the rate of water evaporation from field troughs. For pasture-based systems, the availability of pasture for dairy cow feed affects the DM intake of a cow's diet and thus cow drinking water. In Ireland, the production of fresh pasture is dependent upon a wide range of factors such as soil type, altitude, local climatic conditions and stocking rates (Hanrahan et al., 2017). As the DM feed intake, live weight and sodium intake influences cow drinking water intake, it is feasible that this, along with individual farm effects (leaks, location of water troughs etc.) contributed to the poor prediction accuracy of dairy farm water consumption.

A relationship between electricity and water consumption was found through milk production, the number of total dairy cows and the utilisation of ground water for pre-cooling milk. Increasing milk production and dairy cow numbers resulted in increased electricity and water consumption. Concurrently, utilising gW for pre-cooling milk decreased electricity consumption by 0.11 standard deviations while increasing total water consumption by 0.06 standard deviations when recycled in an open loop system. This result was concurrent with previous research finding that farms which recycled plate cooler water in an open loop system consumed 41% greater magnitudes of unwarranted water within the milking parlour compared to farms which had no pre-cooling (Shine et al., 2018).

The resulting electricity model may be differentiated from that developed by Todde et al. (2017) as: (1) their MLR model predicted annual electricity consumption while this study focused on the prediction of monthly consumption to allow for usage trends throughout the year to be developed. (2) their MLR model included the number of lactating cows as the only input variable. Although this model was found to predict dairy farm electricity consumption to within 11.4% (leave one out cross validation), this model is unable to offer decision support for farmers wishing to optimise their dairy farming infrastructure due to its single variable input. I.e. the only variable which may be optimised is the number of lactating cows.

Water footprinting is a popular method for attributing water consumption to specific processes in the milk production cycle (Murphy et al., 2016; Palhares and Pezzopane, 2015; Ridoutt and Pfister, 2010; Zonderland-Thomassen and Ledgard, 2012). From an Irish perspective, Murphy et al. (2016) utilised 24 dairy farms to calculate a volumetric water footprint of 690 L_w kg^{-1} fat and protein corrected milk. To perform this water footprint analysis, as well as to predict the blue water proportion of freshwater demand (Murphy et al., 2017), the physical metering of on-farm direct water consumption was required. The metering of on-farm water consumption from a borehole or public water supply can be expensive, requiring metering and network equipment to be installed and maintained on each study farm. Thus, this may limit the feasibility of large-scale water footprint analysis. Water footprinting studies such as Ridoutt and Hodges (2017) (75 Australian dairy farms) and Zonderland-Thomassen and Ledgard (2012) (167 New Zealand dairy farms) calculated stock drinking and milking shed water requirements based upon averaged values calculated in benchmark studies (Stewart and Rout, 2007; Victorian Government, 2010). However, an empirical water model for total dairy farm (leak inclusive) water consumption, such as the one developed in this study, may facilitate larger scale water footprinting to be carried out, providing water consumption estimates specific to a particular dairy farm's size and infrastructure.

The water model was developed using a similar variable input approach as Higham et al. (2016). That is, developing the MLR model using 'easily attainable' variables related to milk production, stock, environmental conditions and infrastructural equipment (with this study including inputs related to managerial processes). The benefit of this approach is that MLR input variables can be gathered on a large scale without specialised equipment. However, this study and that of Higham et al. (2016) differs on three levels: (1) Higham et al. predicted daily dairy farm water consumption. Due to the availability of milk production figures and dairy cow numbers on a monthly basis, a monthly prediction of dairy farm electricity and water was carried out in this study. (2) Higham et al. adjusted water consumption using a minimum night flow leak estimation method (Cheung et al., 2010; Tabesh et al., 2005) whereby 26% of stock drinking water was classified as leakage (highly variable between farms). Over 60% of farms in this study manually reported water consumption on a monthly basis, which prevented any adjustment for leaks to be carried out. This may explain the improved prediction results of Higham et al.'s water model ($r = 0.90$) compared to this study ($r = 0.52$).

Both Todde et al. (2017) and Higham et al. (2016) selected input variables based upon their correlation with seen consumption data. More specifically, Todde et al. (2017) selected variables that were significantly (p value < 0.001) associated with electricity consumption for inclusion in the final MLR model. Although a valid methodology, for the final model, their method selected variables that explained the variability of seen data as opposed to unseen data. Thus, their model inputs may not be the optimum model configuration for future unseen data predictions. Similarly, Higham et al. (2016) selected model input variables based upon their ability to explain variability of seen data. Model development was carried out by firstly utilising 31 variables resulting in a model with 38 inputs, made up of continuous, categorical (binary) and squared terms. Variables with the lowest influence (lowest standardised regression coefficient) on total water consumption were iteratively removed to produce a model with a similar R^2 value to the original but with considerably less input variables. Model accuracy was then carried out on unseen data using a leave-one-out cross validation. Albeit a valid methodology, the optimum subset of variables to predict unseen data may not have been selected. In comparison, the methodology presented in this study firstly incorporated a univariate correlation analysis on the E&W datasets to extract high predictive yielding variables similar to Todde et al. (2017) and Higham et al. (2016). However, variables were selected for input to the final MLR models based upon their ability to explain the variability of unseen data. This was carried out through ASR with an embedded k-fold cross validation, whereby the prediction accuracy was calculated through k-fold cross validation for all possible subsets of high predictive yielding variables. This method increased the probability that the optimum variables were selected to calculate MLR coefficients for future predictions of unseen E&W data.

4. Conclusion

This paper presents results from the multiple linear regression (MLR) modelling of Irish dairy farm electricity and water consumption utilising data related to milk production, stock numbers, infrastructural equipment, managerial processes and environmental conditions. The prediction of electricity consumption, based on relative prediction error (RPE), was found to be 47% more accurate than that of water consumption, with RPE values of 26% and 49% for electricity and water, respectively. Milk production, the number of total dairy cows and the utilisation of ground water for pre-cooling milk were found to influence both electricity and water consumption on dairy farms. Increasing milk production and dairy cow numbers resulted in increased electricity and water consumption while utilising ground water for pre-cooling milk decreased electricity consumption by 0.11 standard deviations and increasing total water consumption by 0.06 standard deviations when recycled in an open loop system. Compared to the January, February, November and December period, the electricity model offered a 62% reduced overprediction of electricity consumption between months March and October. Similarly, the water model offered a reduction of 44% in the overprediction during the March–October period when compared to the January, February, November and December period. During the January, February, November and December period, milk production variance between farms increased, which lead to an increase in the variance of electricity and water consumption between farms. The developed MLR models may be used by governing bodies to calculate the impact of Irish dairy farming on resource consumption for international comparisons, marketing Irish dairy products abroad and/or as decision support tools for farmers and/or policy makers to calculate potential impacts of on-farm mitigation practises. Further work will entail the development and analysis of the ability of various machine-learning algorithms to improve the prediction capabilities of dairy farm electricity and water consumption.

Acknowledgements

We would firstly like to acknowledge the 58 farmers who took part in this study for their help and assistance in collecting essential electricity and water consumption data as well as the Irish Cattle Breeding Federation and Met Éireann for the sharing of stock and environmental data used in the development of this study. We would also like to thank Kevin McNamara and Anne Geoghegan of Teagasc Moorepark for their assistance in acquiring and processing of electricity and water consumption measurements and farm infrastructural and managerial information. The Institutes of Technology Ireland (IOTI), Department of Agriculture, Food and the Marine (DAFM) and the Sustainable Energy Authority of Ireland (SEAI) supported this work.

References

Akhil, G., Kang, T., 2013. Comparison of statistical and machine learning methods in modelling of data with multicollinearity. Int. J. Model. Identif. Control 18. http://dx.doi.org/10.1504/IJMIC.2013.053535.

ARTICLE IN PRESS

Baudracco, J., Lopez-Villalobos, N., Holmes, C.W., Comeron, E.A., Macdonald, K.A., Barry, T.N., 2013. E-Dairy: a dynamic and stochastic whole-farm model that predicts biophysical and economic performance of grazing dairy systems. Animal 7, 870–878. http://dx.doi.org/10.1017/S1751731112002376.

Bibby, J., Toutenburg, H., 1977. Prediction and Improved Estimation in Linear Models. Wiley, New York, NY. http://dx.doi.org/10.1002/bimj.197800029.

Brown, C.E., 1998. Coefficient of Variation, Applied Multivariate Statistics in Geohydrology and Related Science. Springer, Berlin Heidelberg doi: 10.1007/978-3-642-80328-4.

Bruinsma, J., Alexandratos, N., 2012. World Agriculture Towards 2030/2050: The 2012 Revision. ESA Work. Pap. No. 12-03. URL http://www.fao.org/docrep/016/ap106e/ap106e.pdf.

Cardot, V., Le Roux, Y., Jurjanz, S., 2008. Drinking behavior of lactating dairy cows and prediction of their water intake. J. Dairy Sci. 91, 2257–2264. http://dx.doi.org/10.3168/jds.2007-0204.

Cheung, P.B., Guilherme, V., Abe, N., Propato, M., 2010. Night flow analysis and modeling for leakage estimation in a water distribution system, 509–513.

DAFM, 2010. (Department of Agriculture, Food and the Marine), Food Harvest 2020 – A Vision for Irish Agri-food and Fisheries . URL https://www.agriculture.gov.ie/media/migration/foodindustrydevelopmenttrademarkets/agri-foodandtheeconomy/foodharvest2020/2020FoodHarvestExeSummary240810.pdf.

DAFM, 2016. (Department of Agriculture, Food and the Marine), FoodWise 2025 – Steps to Success . URL https://www.agriculture.gov.ie/media/migration/foodindustrydevelopmenttrademarkets/agri-foodandtheeconomy/foodwise2025/stepstosuccess2016/FoodWise2025StepstoSuccess2016.pdf.

Daly, J.C., Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, Revised edition. J. Am. Stat. Assoc. Academic Press Inc., New York. http://dx.doi.org/10.2307/2286629.

Edens, W.C., Pordesimo, L.O., Wilhelm, L.R., Burns, R.T., 2003. Energy use analysis of major milking components at a dairy experiment station. Appl. Eng. Agric. 19, 711–716.

Elliot, M., Fairweather, I., Olsen, W., Pampaka, M., 2016. A Dictionary of Social Research Methods, first ed. Oxford University Presshttp://dx.doi.org/10.1093/acref/9780191816826.001.0001.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 2010. A density-based algorithm for discovering clusters. Compr. Chemom. 2, 635–654. http://dx.doi.org/10.1016/B978-044452701-1.00067-3.

Fuentes-Pila, J., DeLorenzo, M.A., Beede, D.K., Staples, C.R., Holter, J.B., 1996. Evaluation of equations based on animal factors to predict intake of lactating Holstein cows. J. Dairy Sci. 79, 1562–1571. http://dx.doi.org/10.3168/jds.S0022-0302(96)76518-9.

Hanrahan, L., Geoghegan, A., O'Donovan, M., Griffith, V., Ruelle, E., Wallace, M., Shalloo, L., 2017. PastureBase Ireland: a grassland decision support system and national database. Comput. Electron. Agric. 136, 193–201. http://dx.doi.org/10.1016/j.compag.2017.01.029.

Herrig, I.M., Böer, S.I., Brennholt, N., Manz, W., 2015. Development of multiple linear regression models as predictive tools for fecal indicator concentrations in a stretch of the lower Lahn River, Germany. Water Res. 85, 148–157. http://dx.doi.org/10.1016/j.watres.2015.08.006.

Higham, C.D., Horne, D., Singh, R., Kuhn-Sherlock, B., Scarsbrook, M.R., 2016. Water use on nonirrigated pasture-based dairy farms: combining detailed monitoring and modeling to set benchmarks. J. Dairy Sci. 100, 1–13. http://dx.doi.org/10.3168/jds.2016-11822.

ICBF, 2016. Irish Cattle Breeding Federation. URL http://www.icbf.com/ (accessed 8.18.16).

Lin, L.I., 2016. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255–268. http://dx.doi.org/10.2307/2532051.

Malathi, A., Rajarajeswari, P., 2014. An efficient enhanced clustering algorithm of information system for law enforcement 8, 144–148

McBride, G., Bland, J.M., Altman, D.G., Lin, L.I., 2005. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. NIWA Client Rep. 45, 307–310. http://dx.doi.org/10.2307/2532051.

Met Éireann, 2017. Met Éireann - Monthly Data. URL http://www.met.ie/climate/monthly-data.asp (accessed 4.4.17).

Meyer, U., Everinghoff, M., Gädeken, D., Flachowsky, G., 2004. Investigations on the water intake of lactating dairy cows. Livest. Prod. Sci. 90, 117–121. http://dx.doi.org/10.1016/j.livprodsci.2004.03.005.

Murgia, L., Todde, G., Caria, M., Pazzona, A., 2013. A partial life cycle assessment approach to evaluate the energy intensity and related greenhouse gas emission in dairy farms. J. Agric. Eng. 44, 186–190. http://dx.doi.org/10.4081/jae.2013. (s1):e37.

Murphy, M.R., Davis, C.L., McCoy, G.C., 1983. Factors affecting water consumption by holstein cows in early lactation. J. Dairy Sci. 66, 35–38. http://dx.doi.org/10.3168/jds.S0022-0302(83)81750-0.

Murphy, M.D., O'Mahony, M.J., Shalloo, L., French, P., Upton, J., 2014. Comparison of modelling techniques for milk-production forecasting. J. Dairy Sci. 97, 3352–3363. http://dx.doi.org/10.3168/jds.2013-7451.

Murphy, E., de Boer, I.J.M., van Middelaar, C.E., Holden, N.M., Shalloo, L., Curran, T.P., Upton, J., 2016. Water footprinting of dairy farming in Ireland. J. Clean. Prod. http://dx.doi.org/10.1016/j.jclepro.2016.07.199.

Murphy, E., De Boer, I.J.M., van Middelaar, C.E., Holden, N., Curran, P., Upton, J., 2017. Predicting fresh water demand on Irish dairy farms using farm data. Clean. Prod. 166, 58–65. http://dx.doi.org/10.1016/j.jclepro.2017.07.240.

Ngo, T. Hoang, 2012. The steps to follow in a multiple regression analysis. In: SAS Institute Inc (Ed.), Proceedings of the SAS Global Forum 2012 Conference. Cary, North Carolina, pp. 1–12.

O'Connor, D., Kean, M., 2014. Future Expansion of the Dairy Industry in Cork : Economic Benefits and Infrastructural Requirements . URL http://mathematics.cit.ie/contentfiles/Dairy Industry_Infrastructure Report Jan27w.pdf.

Palhares, J.C.P., Pezzopane, J.R.M., 2015. Water footprint accounting and scarcity indicators of conventional and organic dairy production systems. J. Clean. Prod. 93, 299–307. http://dx.doi.org/10.1016/j.jclepro.2015.01.035.

Rajaniemi, M., Jokiniemi, T., Alakukku, L., Ahokas, J., 2017. Electric energy consumption of milking process on some finnish dairy farms. Agric. Food Sci. 26, 160–172. http://dx.doi.org/10.23986/afsci.63275.

Ridoutt, B., Hodges, D., 2017. From ISO14046 to water footprint labeling: A case study of indicators applied to milk production in south-eastern Australia. Sci. Total Environ. 599–600, 14–19. http://dx.doi.org/10.1016/j.scitotenv.2017.04.176.

Ridoutt, B.G., Pfister, S., 2010. A revised approach to water footprinting to make transparent the impacts of consumption and production on global freshwater scarcity. Glob. Environ. Change 20, 113–120. http://dx.doi.org/10.1016/j.gloenvcha.2009.08.003.

Rook, A.J., Gill, M., 1990. Prediction of the voluntary intake of grass silages by beef cattle. 1. Linear regression analyses. Anim. Prod. 50, 425–438. http://dx.doi.org/10.1017/S0003356100004918.

Satopää, V., Albrecht, J., Irwin, D., Raghavan, B., 2011. Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In: Proc. - Int. Conf. Distrib. Comput. Syst., pp. 166–171. http://dx.doi.org/10.1109/ICDCSW.2011.20.

Schroeder. D, L., Sjoquist.L, D., Stephan.E, P., 1996. Understanding Regression Analysis: An Introductory Guide. Sage Publications.

Sharma, A.K., Sharma, R.K., Kasana, H.S., 2006. Empirical comparisons of feed-forward connectionist and conventional regression models for prediction of first lactation 305-day milk yield in Karan Fries dairy cows. Neural Comput. Appl. 15, 359–365. http://dx.doi.org/10.1007/s00521-006-0037-y.

Shine, P., Scully, T., Upton, J., Shalloo, L., Murphy, M.D., 2018. Electricity & direct water consumption on Irish pasture based dairy farms: a statistical analysis. Appl. Energy 210, 529–537. http://dx.doi.org/10.1016/j.apenergy.2017.07.029.

Stewart, G., Rout, R., 2007. Reasonable Stock Water Requirements: Guidelines for Resource Consent Applications . URL https://www.boprc.govt.nz/media/470831/reasonable-stock-water-requirements-guidelines-horizons.pdf.

Tabesh, M., Yekta, A., Hossein, A., 2005. Assessment of real losses in potable water distribution systems: some recent developments. Water Sci. Technol. Water Supply 5, 33–40. http://dx.doi.org/10.1007/s11269-008-9284-2.

Todde, G., Murgia, L., Caria, M., Pazzona, A., 2017. Dairy Energy Prediction (DEP) model: a tool for predicting energy use and related emissions and costs in dairy farms. Comput. Electron. Agric. 135, 216–221. http://dx.doi.org/10.1016/j.compag.2017.02.014.

Upton, J., Humphreys, J., Groot Koerkamp, P.W.G., French, P., Dillon, P., De Boer, I.J.M., 2013. Energy demand on dairy farms in Ireland. J. Dairy Sci. 96, 6489–6498. http://dx.doi.org/10.3168/jds.2013-6874.

Upton, J., Murphy, M., Shalloo, L., Groot Koerkamp, P.W.G., De Boer, I.J.M., 2014. A mechanistic model for electricity consumption on dairy farms: definition, validation, and demonstration. J. Dairy Sci. 97, 4973–4984. http://dx.doi.org/10.3168/jds.2014-8015.

Victorian Government, 2010. Dairy Shed Water Use in Victoria: 2009 Analysis. URL http://agriculture.vic.gov.au/_data/assets/pdf_file/0003/197085/Dairy-Shed-Water-Use-in-Victoria-2009-Analysis.pdf.

Zhang, F., Murphy, M.D., Shalloo, L., Ruelle, E., Upton, J., 2016. An automatic model configuration and optimization system for milk production forecasting. Comput. Electron. Agric. http://dx.doi.org/10.1016/j.compag.2016.08.016.

Zonderland-Thomassen, M.A., Ledgard, S.F., 2012. Water footprinting – a comparison of methods using New Zealand dairy farming as a case study. Agric. Syst. 110, 30–40. http://dx.doi.org/10.1016/j.agsy.2012.03.006.