Check for updates

Modelling methane emissions from pig manure using statistical and machine learning methods

Jayanta Kumar Basak[1,2] · Elanchezhian Arulmozhi[3] · Byeong Eun Moon[1] · Anil Bhujel[3] · Hyeon Tae Kim[3]

Abstract

Manure production and its management in the livestock sector have been increasingly receiving global attention due to its contribution to generating greenhouse gases, especially methane (CH_4). This study was conducted to quantify and characterize daily manure including its moisture, dry matter (DM), ash, volatile solid (VS) contents, and model CH_4 production rate as a function of feed intake and mass of pigs. Two statistical (multiple linear regression and polynomial regression) and three machine learning algorithms (ridge regression, random forest regression, and artificial neural network) were employed to predict CH_4 emission. The result showed body mass ranged from 60 to 90 kg pig produced around 4.78 kg of manure per day consisting of 67% moisture content and 33% DM. The manure's ash content was 28% DM (0.45 kg pig^{-1} day^{-1}), while the VS was 72% DM (1.21 kg pig^{-1} day^{-1}). Moreover, the average CH_4 production rate was estimated as 0.018 kg pig^{-1} day^{-1} which was lower than IPCC's (2006) recommended value for Oceania, Western Europe, and even North America regions. The current study found that the performance of the ridge regression was comparatively better, where the model with a coefficient of determination (R^2) greater than 90% was suitable for describing the relationship between the explanatory (feed intake and mass of pigs) and the response (CH_4 emissions) variables. Further research may be conducted to improve this model's prediction accuracy, providing a wider range of diets and management conditions.

Keywords Emission · Manure · Methane · Model · Pig

Introduction

Worldwide agriculture and livestock sectors contribute significantly to the anthropogenic emissions of greenhouse gases (GHGs), predominantly methane (CH_4) (Riaño and García-González 2015; Ngwabie et al. 2018). Globally, about 18% of anthropogenic emissions occur from livestock production whereas 37% of anthropogenic CH_4 generates only from ruminant enteric fermentation and manure management (Steinfeld et al. 2006). Among the livestock animals, pig contributes about 13% of the total emissions of GHGs which is the second-highest source of GHGs emission in the livestock sector (FAO 2011). In recent years, the total pig production in Korea has rapidly increased due to the changes in dietary life (Lu et al. 2008; Wang et al. 2010) and hence, CH_4 emissions are significant from pig farming. Ji and Park (2012) noted that the annual growth rate of CH_4 emission from manure management in Korea was 2.6% from 1990 to 2009, which has a substantial contribution to GHGs on climate change. Therefore, CH_4 emission modelling can act as a preliminary step for controlling the mechanisms of GHGs. It gives a clear view of the CH_4 emission rate from pig manure based on the quantity of feed intake and age of pig, which might be helpful to reduce the CH_4 emission rate in pig barns. Within this context, this study was conducted to characterize and quantify pig manure for methane emission modelling.

Methane emission from pig's manure is mainly affected by environmental conditions, manure properties (e.g., moisture content and pH), manure management practices, and the portion of manure anaerobically decomposed (Liang et al. 2005; Cortus et al. 2012; Hetchler et al. 2015). Moreover, the

✉ Hyeon Tae Kim
  bioani@gnu.ac.kr

1 Institute of Smart Farm, Gyeongsang National University, Jinju 52828, Korea

2 Department of Environmental Science and Disaster Management, Noakhali Science and Technology University, Noakhali 3814, Bangladesh

3 Department of Bio-Systems Engineering, Institute of Smart Farm, Gyeongsang National University, Jinju 52828, Korea

manure properties primarily depend upon the nutrient value of the feed given to the pig and the age of the pig. It is noted that the nitrogen excretion rate decreased as a percent of the bodyweight of pigs when the concentration of protein in the diet was reduced (Ogejo 2019). Since different factors influence in generating methane, it is essential to select appropriate input variables, which are informative for methane emission modelling. Moreover, to maximize the accuracy and minimize the error of model estimation, it is important to develop a precise model with the best architecture. Recently, a number of emission models such as regression equations, manure-DeNitrification DeComposition (DNDC), empirical model of CH_4 emissions have been used to estimate the CH_4 emission rate (Olesen et al. 2006; Li et al. 2012; Petersen et al. 2016; Ngwabie et al. 2018). For example, the relationship of CH_4 production rate as a function of pigs' mass was described through an exponential equation with coefficients of determination (R^2) > 88% (Ngwabie et al. 2018). Manure-DNDC modeled results showed that crop cultivation and lagoon coverage with feed quality changes could reduce GHGs emissions by 30% at the farm scale (Li et al. 2012). These CH_4 emission models were developed based on various methodologies including emission factors, empirical equations, and process-oriented mechanisms. The Intergovernmental Panel on Climate Change (IPCC 2006) also prepared guidelines based on the amount of volatile solids in manure for estimating CH_4 emission in country-specific (tier 2 approach). Several studies have been conducted on the tier 2 approach to calculate country-specific CH_4 emissions rate (ANIR 2012; Environment-Canada 2012; Du Toit et al. 2013; Ngwabie et al. 2018).

Sometimes, it seems challenging to figure out the best models among the available mathematical and statistical techniques. Considering the previous research, the current study evaluates the performance of five models i.e., multiple linear regression (MLR), polynomial regression (PR), ridge regression (RR), random forest regression (RFR), and artificial neural network (ANN) on CH_4 emission modeling. Regression-based models is a very common term for a group of diverse statistical methods that are widely applied in various contexts in different fields (Seuront 2010; Cranford et al. 2011; Hirst 2012; Carey et al. 2013; Beninger and Boldina 2014; Ngwabie et al. 2018; Basak et al. 2019; Ekine-Dzivenu et al. 2020; Font-i-Furnols et al. 2021). In contrast, ML models have been utilized to overcome non-linear data processing and accomplish better prediction accuracy efficiently (Joharestani et al. 2019; Shin et al. 2000). However, a limited study was found to investigate ANN and RFR algorithms for environmental pollution modelling (Rybarczyk and Zalakeviciute 2018; Shahriar et al. 2020).

Estimating CH_4 emission from pig manure requires an elaborate process including the establishment of the experimental setup, maintaining environmental conditions, dietary composition, quantity of feed intake, etc. Additionally, the CH_4 emission rate varies considerably from year to year and within the same experimental pig barn, making it even more challenging to measure. In that sense, regression and ML techniques might be useful in predicting CH_4 emission. These techniques reduce sampling efforts and cost and increase precision where samples are difficult to handle (Basak et al. 2019). As such, the objectives of the research are to characterize and quantify the daily manure production rates for its moisture, dry matter (DM), ash, and volatile solid (VS) daily excretion rates and finally to model CH_4 production rates as a function of the feed intake and mass of pigs using regression and ML methods.

Materials and methods

Animal resources, experimental design, and data collection

The research methods and procedures were approved by the ethics and animal experimentation committee of the Gyeongsang National University (certification# GNU-150508-R0029). Two independent experiments were performed from 1st September to 1st December in 2019 and 2020 with six 2-month-old Yorkshire breed pigs in three experimental pig barns in Smart Farm Systems Laboratory at Gyeongsang National University. Three pig barns had slatted floors, with sidewalls made of galvanized steel and plywood, and an expanded polystyrene roof. In every experimental period, similar sizes of the pigs (ages and weights) were studied with three concentrated diets (Table 1). Similar sizes of pigs were kept in all barns, so that, it could be expected that manure production rates were identical. A two-week observation phase was implemented before beginning the experiment to define the best data measuring conditions. In both of the experimental periods, the Health-Feed Gatdon3 (Daejoo Inc., Seoul, Republic of Korea) concentrated diet was provided to the pigs allotted to pig barn 1 (PB1), G Max Care (Growing pigs) (Nonghyup Feed Co., Ltd., Seoul, Republic of Korea feed) for pig barn 2 (PB2) and Growing Pigs Late Feed 10 (Nonghyup Feed Co., Ltd., Seoul, Republic of Korea) for pig barn 3 (PB3). An equal amount of feed was provided two times in a day, at 10.00 am and 17.00 and the amounts of feed intake was estimated from the daily recorded of feed offered and leftovers of each pig. The load cell was used to determine the pigs' mass by averaging weights measured two times in a day. Moreover, drinkers and feeders were installed in the barns to restrain with halters for feeding and drinking (Fig. 1).

Under each pen, polythene sheets were put to capture urine and faecal matter to calculate the rate of manure production of various types of diets. For covering the full

Table 1 Three concentrate diets and their ingredients

Ingredient	Feed type and ingredient		
	Health-Feed Gatdon3	G Max Care (growing pigs)	Growing Pigs Late Feed 10
Crude protein	Below 20.0%	Below 18.0%	Below 18.0%
Crude fat	More than 6.5%	More than 3.0%	More than 4.5%
Crude fiber	Below 5.0%	Below 9.0%	Below 10.0%
Crude ash	Below 8.0%	Below 10.0%	Below 8.0%
Calcium	More than 0.65%	More than 0.45%	More than 0.5%
Phosphorus	Below 1.0%	Below 1.5%	Below 1.2%
Lysine	More than 1.25%	More than 0.9%	More than 0.9%
Digestible crude protein (DCP)	More than 16.5%	More than 12.0%	More than 12.0%
Digestible energy (DE)	3,600 kcal kg^{-1}	3,350 kcal kg^{-1}	3,500 kcal kg^{-1}
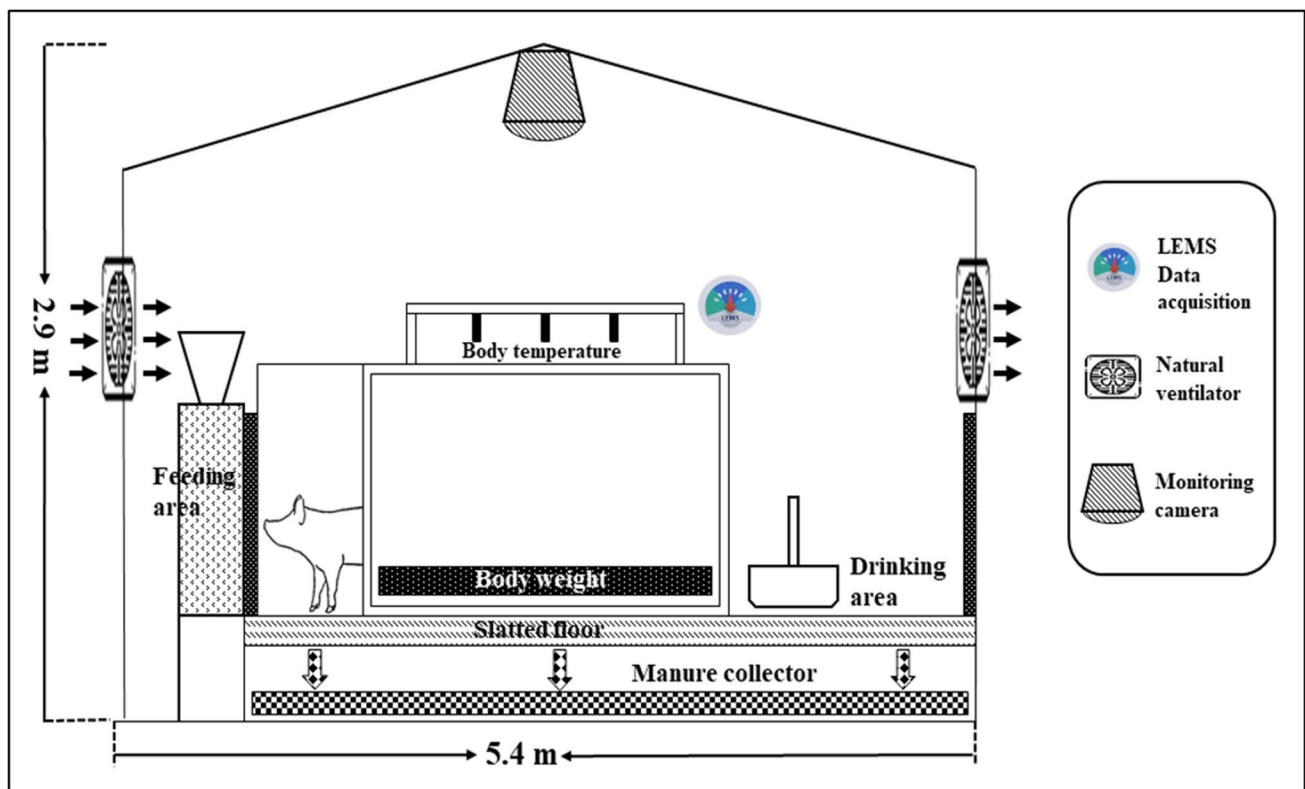
Fig. 1 Layout of the experimental pig's barn

surface area underneath the pen, slightly larger polythene sheets were used. It was raised to stop urine and faecal matter from escaping from the surface region of the manure. The polythenes were placed every morning before feeding and removed 24 h later (Ngwabie et al. 2018). The manure collected in each pig barn was weighed to measure all the pigs' production rates over 24 h. To estimate the manure produced by a pig in a day (kg manure pig^{-1} day^{-1}), the total amount of manure was divided by the number of pigs. The manure obtained from each pig barn, a small portion was

subsequently used in the laboratory to analyze its moisture content, DM, ash, and VS content. Moreover, the pig's body surface temperature (PBT) was measured using infrared sensors (IR sensor, model-MI3, Raytek Corporation, CA, USA) at 10.00 am and 17.00 two times a day. Temperature, humidity, and carbon dioxide data inside and outside of the pig barns were collected by using Livestock environment management systems (LEMS, AgriRoboTech Co., Ltd, Republic of Korea) and weather sensors (MetPRO, Producer: Campbell Scientific, USA), respectively.

Measurement of manure parameters

The pH level in each manure sample was measured using a portable pH meter (HP9010, Trans Instruments (S) Pte Ltd, Singapore) in a day. The DM and VS in manure were calculated according to the method 1648 of the U.S. Environmental Protection Agency (Telliard 2001). The overall procedure was conducted in the following steps. (i) A portion of manure in each sample was weighed using an electronic mass balance (model-FX-300iWP, A&D Company Limited, Tokyo, Japan) to obtain its wet mass (M_w). (ii) It was oven-dried at 105 °C for 12 h using 5E-DHG6340 drying oven (Shelves for 5E-DHG6310: 2 Layers, Changsha Kaiyuan Instruments Co., Ltd, China), after which it was weighed to obtain the dry mass (M_d). (iii) Eqs. (1), (2), and (3) were applied to calculate the percentage of DM and the daily DM excreted by each pig as well as the moisture content of the manure. (iv) In order to calculate the VS content from each oven-dried sample, a part of the sample was weighed in a crucible of known mass M_c to obtain its mass (M_1). The crucibles with samples were then placed in a muffle furnace (Digital Muffle Furnace 14 Lit "FX-14" 1,000 °C, Korea) and heated at 450 °C for 4 h. After cooling the sample in the furnace, the remaining part of the heated sample (ash) and crucibles were weighed again to obtain the mass M_2. The VS and ash as a percentage of the DM were calculated according to Eqs. (4) and (5). (v) Finally, Eq. (6) was used to calculate the daily VS excreted per pig.

$$DM = \frac{M_d}{M_w} \times 100 \tag{1}$$

$$DM_{pig} = \frac{DM\%}{100} \times M_{Pig} \tag{2}$$

$$M_c(\%) = \frac{M_w - M_d}{M_w} \times 100 \tag{3}$$

$$VS_{DM}(\%) = \frac{(M_1 - M_c) - (M_2 - M_c)}{(M_1 - M_c)} \tag{4}$$

$$Ash_{DM}(\%) = 100 - VS_{DM}(\%) \tag{5}$$

$$VS_{pig} = \frac{VS_{DM}(\%)}{100} \times DM_{pig} \tag{6}$$

Measurement of methane production rate

The CH_4 production rate from pig manure was calculated using the IPCC tier 2 approach shown in Eq. (7) (IPCC 2006). Several studies have used the same approach to calculate the CH_4 production rate (ANIR 2012; Du Toit et al. 2013; Shin et al. 2016; Ngwabie et al. 2018). According to the IPCC tier 2 approach, the model requires country-specific input values for VS excreted from manure, the maximum CH_4 producing capacity (B_0) for pig manure, and CH_4 conversion factor (MCF). Here, MCF indicates that the percentage of VS in manure converts to CH_4 compared to the theoretical maximum.

$$EF = VS \times B_0 \times MCF \times 0.67 \tag{7}$$

where EF is the CH_4 production rate (kg CH_4 pig^{-1} day^{-1}). VS is the daily volatile solid excreted (kg VS pig^{-1} day^{-1}) from manure in the present experiment. The B_0 value primarily depends on the type of diet (Kumar et al. 2014). It is reported that feeding a high forage diet leads to more methane emissions compared to a concentrated diet (Won et al. 2014). In Korea, pig-breeding circumstances prefer a more concentrated diet than forage (Won et al. 2014). Thus, in the present work, B_0 value of 0.0579 m^3 CH_4 Kg^{-1} of VS excreted (Won et al. 2014) and MCF value of 0.39 (Park et al. 2006) were used to calculate the CH_4 production rate, while 0.67 is the conversion factor from m^3 CH_4 to kg CH_4.

Data analysis and model development

Data obtained through two experimental periods from the three pig barns were used to develop statistical and machine learning methods for CH_4 emission modelling. During the model preparing stage, the mass of pig (MP), age, and feed intake (FI) are considered input variables. However, high multicollinearity was examined between body weight and age values (correlation coefficient (r) = 0.93); thus, MP and FI were selected as inputs variables in the present study. The Z-score data normalization technique (Eq. 8) was used to keep values within a scale applied across all numeric columns used in the model.

$$Z = \frac{x - \mu}{\sigma} \tag{8}$$

where z is the standard score; x is the value in the data set; μ is the mean of all values in the data set; and σ is the standard deviation.

Recently, many statistical and machine learning (ML) methods are being utilized for both prediction and inference in different research fields. In this study, two statistical models, i.e., multiple linear regression (MLR) and polynomial regression (PR), and three ML algorithms, i.e., ridge regression (RR), random forest regression (RFR), and feed forward-back propagation (FFBP), in ANN were evaluated based on how well these algorithms predicted CH_4 emission

Table 2 Year wise compositions diets, pig barns, datasets and number of data

Year	Diet	Pig barn	Dataset	Data
2019 2020	Health-Feed Gatdon3	PB1	F1	184
2019 2020	G Max Care (Growing pigs)	PB2	F2	184
2019 2020	Growing Pigs Late Feed 10	PB3	F3	184
2019, 2020	F1, F2, F3	PB1, PB2, PB3	FC	552

from the four datasets presented in Table 2. One of the main reasons for using statistical models is that they help for intuitive visualizations of data that aid in identifying relationships between variables and making predictions. ML models, on the other hand, focus on prediction, employing general-purpose learning algorithms to uncover patterns in often complex and unwieldy data. Moreover, the rationale for using these ML models was its specialty to predict non-linear interactions between the experimental and predictor variables. In the following sections of methodology, the statistical and ML approaches used for predicting CH_4 emission are briefly discussed.

Multiple linear regression

Multiple linear regression (MLR) has the ability to model explanatory (MP and FI) and response variables (CH_4 emission rate) more simply and comprehensively (Tabachnick and Fidell 2001). In the present study, the MLR model was developed according to the Eq. (9) (Darlington and Hayes 2016):

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots + \beta_n X_n + \varepsilon \tag{9}$$

where Y_i is the CH_4 emission rate, β_0–β_n are the coefficients of regression, X_1–X_n are the input variables, and ε is the error associated with the ith observation.

Polynomial regression

The polynomial regression (PR) is also a form of regression in which a non-linear relationship between the explanatory and response variables is modeled as a degree polynomial. Therefore, PR is considered to be a particular case of the MLR model (Ostertagová, 2012). The general form of a complete second-degree PR model with two independent variables X_1 and X_2 as shown Eq. (10).

$$\mu_{y|(X1,X2)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2 + \varepsilon \tag{10}$$

where $\mu_{y|(X1, X2)}$ is the true mean response for the two independent variables (MP and FI), β (0, 1, 2, 3, 4, 5) is the model parameters, and X (1, 2) is the independent variable. After experimenting on different degrees of polynomials (order = 2, 3, 4, 5), the present work decided to use 3-degree polynomial regression due to its better performance compared to others.

Ridge regression

In the present research work, a penalty-based regression procedure (i.e., ridge regression (RR)) was also used to model CH_4 emission. The RR has been widely utilized to measure many characteristics of a single sample simultaneously (Tibshirani 1996; McDonald 2009; Ransom et al. 2019; Wieringen 2020). It is a continuous form of shrinkage in which the residual sum of squares is minimized as each parameter's coefficient is adjusted near to zero, thus reducing the importance or influence of any particular parameter (Hoerl and Kennard 2000; Ransom et al. 2019). RR equation is also very close to least square, but the minimization equation is slightly different, as shown in Eq. (11). Specifically, it is as follows.

$$L_{ridge}(\beta^\wedge) = \sum_{i=1}^{n} (y_i - x_i' \beta^\wedge)^2 + \lambda \sum_{j=1}^{m} (\beta^\wedge)_j^2 = \|y - X_B^\wedge\|^2 + \lambda \|^\wedge_B\|^2 \tag{11}$$

where $\|y - X_B^\wedge\|^2$ is called the sum of the squares of all coefficients (RSS) and it is also denoted as a loss function, and the λ parameter is the regularization penalty.

Random forest regression (RFR)

In recent times, random forest regression (RFR) is considered one of the most effective machine learning algorithms (Biau and Scornet 2016). It has been applied to a wide range of learning tasks, but most prominently to classification and regression (Biau et al. 2008; Biau 2012; Denil et al. 2014; Ransom et al. 2019; Shahriar et al. 2020). RFR model is an ensemble of trees, where the construction of each tree is made randomly. After building an ensemble of trees, the RFR model makes predictions by averaging the prediction of an individual tree. Even for very high-dimensional problems, RFR often makes accurate and robust predictions (Biau 2012). The random forest estimator associated with the tree collection is defined by the Eq. (12).

$$\tilde{\eta}_n \mathbb{V}_T(x) := \frac{1}{T} \sum_{j=1}^{T} \tilde{\eta}_n, Aj(x) = \frac{1}{T} \sum_{j=1}^{T} \frac{1}{N(Aj(x))} \sum_{i=1}^{n} Y_i 1_{X_i \in Aj(x)} \tag{12}$$

A more comprehensive introduction to random forest algorithm was reported by Friedman et al. (2009) and Biau and Scornet (2016).

Artificial neural network

The artificial neural network (ANN) models consist of interconnecting artificial neurons by transferring signals to another, along with weighted connections (Feng et al. 2015). The input, hidden, and output layers are needed to create an ANN topology. Every input value is regarded as a neuron in the input layer. All input values are weighted randomly at first, and then, the weighted values are processed into the hidden layers. In the hidden layers, every neuron produces output values for the success of the ANN model. ANN models are widely used to analyse non-linear data (Shin et al. 2000). Different architectures of ANN, like feed forward-back propagation neural network (Basak et al. 2020), adaptive logic network (Qu et al. 2001), radial basis function network (Boilot et al. 2002), self-organizing map network (Sinesio et al. 2000), time-delay neural network (Zhang et al. 2003), and hybrid Bi-GRU-ARIMA model (PAHM) (Jaihuni et al. 2020) have been applied to analyze data in a number of studies. After experimenting on several multilayer perceptrons (MLP) structures with neurons and three transfer functions (Log-sigmoid, linear transfer function (purelin), and Tansigmoid), the study decided to employ FFBP neural network, gradient descent weight and bias learning function, one hidden layer and log-sigmoid transfer function. The output of an ANN network was noted by Hydrology (2000).

$$y_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j f\left(\sum_{i=1}^{m} \beta_{ij} y_{t-1} + \beta_{0j} \right) + \varepsilon_t \tag{13}$$

where y_t is the network output (pig's body temperature), n is the number of hidden nodes, m is the number of input nodes, f is the transfer function, β_{ij} {i = 1, 2,…, m; j = 0, 1,…, n} are the weights from the input to hidden nodes, α_j {j = 0, 1, …, n} are the vectors of weights from the hidden to the output nodes, and α_0 and β_{0j} denote the weights of arcs leading from the bias terms.

Application methodology and performance metrics

All the statistical and ML models were developed using open-source libraries under the Python (Python 3.7.0) environment in the present work. Python is a high-level, interpreted programming language that can be used for different uses, including for scientific purposes (Tran et al. 2020). In the Python platform, different libraries such as NumPy (Van Der Walt et al. 2011), Pandas (McKinney 2010), and Matplotlib (Hunter 2007) were used for processing, manipulating, and visualizing data. In the current study, 70% of the data was selected as the training set, while the remaining 30% was used as the testing dataset. The performance of the models was evaluated on the basis of the two statistical

quality parameters, i.e., root mean square error (RMSE) (Eq. 14) and coefficient of determination (R^2) (Eq. 15).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n}} \tag{14}$$

$$R^2 = \frac{\sum_{i=1}^{n}(O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2} \times \sqrt{\sum_{i=1}^{n}(P_i - \overline{P})^2}} \tag{15}$$

where n is the number of data, O_i is the observed values, P_i is the predicted values, and the bar denotes the mean of the variable. All statistical calculations in this study were performed with Statistical Package for the Social Sciences (IBM SPSS Statistics 22.0.0.0, NY, USA) and Origin Pro 9.5.5 (OriginLab, Northampton, MA, USA).

Results and discussions

Environmental data measurement

Manure temperature was measured from the collected samples ranging from 25.6 to 31.7 °C, 24.3 to 30.8 °C, and 26.1 to 32.4 °C at pig barns 1, 2, and 3, respectively. Variations of air temperature, humidity, and pig's body temperature of the three pig's barns for both experimental periods are shown in Fig. 2. In summary, the air temperature in the PB1 ranged from 9.4 to 31.6 °C, 9.1 to 33.3 °C for PB2, and 9.3 to 32.6 °C for PB3 during the two experimental periods in 2019 and 2020. The average pig's body temperatures for PB1, PB2, and PB3 were 32.28 ± 2.66 °C, 32.29 ± 2.87 °C, and 32.18 ± 2.53 °C, respectively in 2019, and the corresponding average for 2020 were 32.53 ± 2.63 °C, 32.01 ± 2.63 °C, and 31.95 ± 2.54 °C, respectively. In order to have a good understanding of the relationship of ambient environmental parameters and the body temperature of pigs, the result was described in our earlier study (Basak et al. 2020).

Estimation of methane production rate

The first step in this section in addressing this question was to find out if there was a significant statistical difference between the three concentrated diets and the manure production rate. An analysis of variance using the three types of concentrated diets and manure production resulted in an insignificant two-way interaction ($p = 0.85$). Manure production rates, moisture DM ratios, ash, and VS production rates according to the diets are presented in Table 3. It is noteworthy that manure production rates, DM, and VS production from manure increased with the growing of pigs' mass and feed intake. Though the barns had different diets,
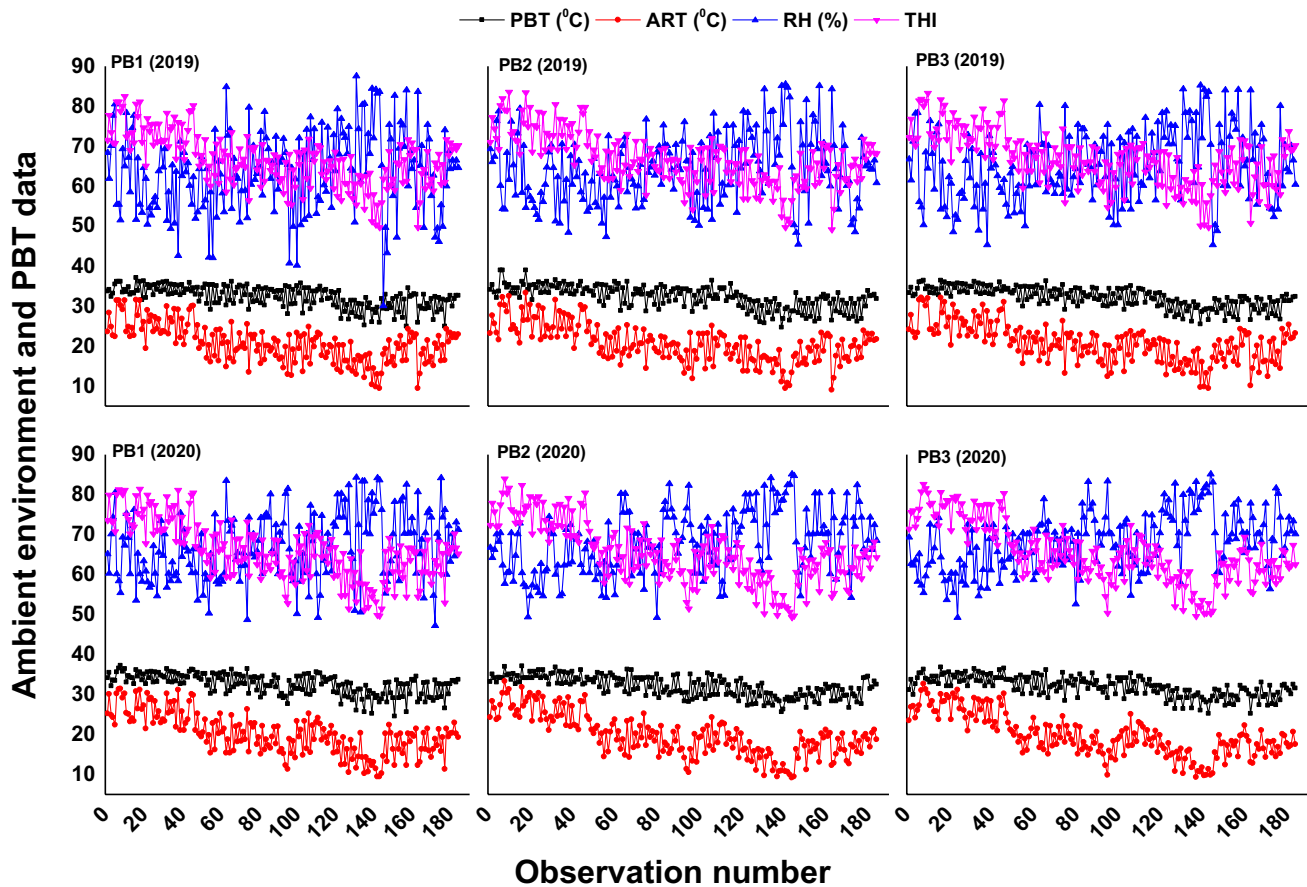
Fig. 2 Variations of air temperature, humidity of the three pig's barn, and pig's body temperature in 2019 and 2020 (data were collected at 10.00 am and 17.00 pm two times a day). ART represents the air room temperature, RH represents the room relative humidity and THI denotes the temperature-humidity index in pig's barn

therefore, comparable DM and VS data were measured. The study result showed that manure DM excretion rates were 1.95 ± 1.10, 1.94 ± 1.13 and 1.86 ± 1.12 kg pig^{-1} day^{-1} for PB1, PB2, and PB3, respectively and manure VS concentrations were 1.42 ± 0.80, 1.41 ± 0.84 and 1.35 ± 0.82 kg pig^{-1} day^{-1} for barns PB1, PB2, and PB3, respectively in 2019. It should, however, be noted that there really was no significant effect of the concentrated diets on varying the production rate of DM ($p = 0.82$) and VS ($p = 0.88$). There was not a statistically significant difference between the diets and DM ($p = 0.82$) and diets and VS concentration rates ($p = 0.88$). Combining the results of the three pig's barns showed that with body mass ranging from 60 to 90 kg, a pig produced around 4.78 ± 1.21 kg of manure per day consisting of $66.51 \pm 4.46\%$ moisture content and $33.49 \pm 4.46\%$ DM. The manure's ash content was $28.74 \pm 4.44\%$ DM (0.46 ± 0.16 kg pig^{-1} day^{-1}), while the VS was $71.26 \pm 4.44\%$ DM (1.13 ± 0.32 kg pig^{-1} day^{-1}). Likewise, manure characteristics have been reported in some studies (Hamilton et al. 1997; IPCC 2006; Won et al. 2014; Dennehy et al. 2017; Shin et al. 2017; Ngwabie et al. 2018).

According to Ngwabie et al. (2018), a 50 kg pig produced approximately 3 kg of manure per day, of which 2.09 kg (~ 70%) was the moisture content and 0.91 kg (~ 30%) was the DM content. It has been widely reported that the diet ratio is crucial in DM and VS production. Another study on corn-based ration showed that pigs of 23–79 kg produced 2.7–3.6 kg of manure per day, of which the VS was 0.24–0.33 kg day^{-1} as excreted (Hamilton et al. 1997; Chastain et al. 1999). A similar report by Chastain et al. (1999) stated that a 60 kg pig produced about 5 kg of manure per day of which 0.51 kg is the VS.

The average CH$_4$ production in all three barns is presented in Table 3, as seen the mean CH$_4$ emission rates were in PB1: 0.021 ± 0.012 kg pig^{-1} day^{-1}, PB2: 0.020 ± 0.013 kg pig^{-1} day^{-1}, and the PB3: 0.019 ± 0.012 kg pig^{-1} day^{-1}. Combining the results from the three barns, the average CH$_4$ production rate was 0.020 ± 0.012 kg pig^{-1} day^{-1}. Figure 3 shows the relationships between the mass of pigs, feed intake, and manure production from pigs with three concentrated diets (F1, F2, and F3) in 2019 and 2020. The modelled CH$_4$ production ranged from 7.20 to 7.70 kg pig^{-1} year^{-1}

Table 3 Feed intake, manure production, and its characterization from pigs of different masses and diets. The values in italics indicate the average results in a year among the diets and bold values indicate the average results in 2019 and 2020

| Year | Feed type | Average mass of pig (kg pig⁻¹) | Average feed intake (kg pig⁻¹ day⁻¹) | Manure characteristics | | | | | | | | CH₄ production |
				Production (kg pig⁻¹ day⁻¹)	Moisture (%)	DM (%)	DM (kg pig⁻¹ day⁻¹)	VS (%DM)	Ash (%DM)	VS (kg pig⁻¹ day⁻¹)	pH	kg (CH₄ pig⁻¹ day⁻¹)
2019	F1	$30 < MP \leq 60$	2.15 ± 0.42^a	1.90 ± 0.60^a	64.34 ± 4.89^a	35.66 ± 4.89^a	0.68 ± 0.24^a	74.20 ± 4.31^a	25.80 ± 4.31^a	0.51 ± 0.18^a	6.55 ± 0.44^a	0.008 ± 0.003^a
		$60 < MP \leq 90$	3.41 ± 0.37^b	4.83 ± 1.30^b	65.25 ± 4.26^a	34.75 ± 4.26^a	1.67 ± 0.46^b	72.83 ± 4.42^a	27.16 ± 4.42^a	1.21 ± 0.33^b	6.66 ± 0.54^a	0.018 ± 0.005^b
		$90 < MP \leq 120$	4.47 ± 0.10^c	8.98 ± 1.06^c	66.86 ± 3.45^a	33.14 ± 3.45^a	2.97 ± 0.38^c	70.57 ± 3.94^a	29.43 ± 3.94^a	2.37 ± 0.31^c	6.80 ± 0.58^a	0.033 ± 0.005^c
		$120 < MP \leq 130$	4.49 ± 0.04^c	10.61 ± 0.77^d	65.81 ± 3.89^a	34.19 ± 3.89^a	3.63 ± 0.51^d	72.02 ± 4.20^a	27.98 ± 4.20^a	2.42 ± 0.43^d	6.46 ± 0.56^a	0.039 ± 0.006^d
	Mean±sd	130	3.46 ± 1.00	5.74 ± 3.31	65.52 ± 4.25	34.48 ± 4.25	1.95 ± 1.10	73.24 ± 4.22	26.76 ± 4.22	1.42 ± 0.80	6.65 ± 0.54	0.021 ± 0.012
	F2	$30 < MP \leq 60$	2.30 ± 0.05^a	2.08 ± 0.69^a	64.79 ± 6.79^a	35.21 ± 6.79^a	0.74 ± 0.31^a	73.93 ± 4.16^a	26.07 ± 4.16^a	0.55 ± 0.23^a	6.73 ± 0.52^a	0.008 ± 0.004^a
		$60 < MP \leq 90$	3.52 ± 0.35^b	5.29 ± 1.05^b	66.15 ± 4.33^a	33.85 ± 4.33^a	1.79 ± 0.44^b	72.46 ± 4.93^a	27.54 ± 4.93^a	1.29 ± 0.30^b	6.54 ± 0.58^a	0.019 ± 0.004^b
		$90 < MP \leq 120$	4.47 ± 0.09^c	8.93 ± 1.08^c	65.90 ± 4.16^a	34.10 ± 4.16^a	3.05 ± 0.55^c	72.32 ± 4.53^a	27.68 ± 4.53^a	2.21 ± 0.31^c	6.79 ± 0.43^a	0.031 ± 0.005^c
		$120 < MP \leq 127$	4.49 ± 0.07^c	10.66 ± 0.66^d	65.56 ± 5.43^a	34.44 ± 5.43^a	3.66 ± 0.45^d	73.60 ± 4.13^a	26.40 ± 4.13^a	2.70 ± 0.43^d	6.51 ± 0.50^a	0.040 ± 0.005^d
	Mean±sd	127	3.45 ± 0.99	5.70 ± 3.23	65.56 ± 5.29	34.44 ± 5.29	1.94 ± 1.13	73.00 ± 4.49	27.00 ± 4.49	1.41 ± 0.84	6.68 ± 0.51	0.020 ± 0.013
	F3	$30 < MP \leq 60$	2.22 ± 0.04^a	1.90 ± 0.70^a	65.40 ± 4.20^a	34.60 ± 4.20^a	0.67 ± 0.27^a	70.80 ± 5.50^a	29.20 ± 5.50^a	0.48 ± 0.19^a	6.67 ± 0.54^a	0.007 ± 0.003^a
		$60 < MP \leq 90$	3.44 ± 0.04^b	4.83 ± 1.29^b	66.90 ± 4.77^a	33.10 ± 4.77^a	1.60 ± 0.51^b	73.05 ± 4.25^a	26.95 ± 4.25^a	1.16 ± 0.34^b	6.73 ± 0.60^a	0.018 ± 0.005^b
		$90 < MP \leq 120$	4.45 ± 0.14^c	8.86 ± 1.05^c	67.13 ± 5.89^a	32.87 ± 5.89^a	8.86 ± 0.70^c	72.70 ± 4.37^a	27.30 ± 4.37^a	2.07 ± 0.50^c	6.87 ± 0.53^a	0.031 ± 0.008^c
		$120 < MP \leq 129$	4.49 ± 0.08^c	10.55 ± 0.75^d	67.00 ± 5.13^a	33.00 ± 5.13^a	3.46 ± 0.37^d	73.91 ± 2.99^a	26.09 ± 2.99^a	2.55 ± 0.25^d	6.75 ± 0.47^a	0.039 ± 0.004^d
	Mean±sd	129	3.46 ± 0.99	5.60 ± 3.26	66.49 ± 4.98	33.51 ± 4.98	1.86 ± 1.12	72.32 ± 4.68	27.68 ± 4.68	1.35 ± 0.82	6.75 ± 0.53	0.019 ± 0.012
	Mean±sd	*130*	*3.46 ± 1.01*	*5.67 ± 3.26*	*65.86 ± 4.86*	*34.14 ± 4.86*	*1.92 ± 1.11*	*72.85 ± 4.46*	*27.15 ± 4.46*	*1.39 ± 0.82*	*6.70 ± 0.53*	*0.021 ± 0.012*
2020	F1	$30 < MP \leq 60$	2.15 ± 0.42^a	1.96 ± 0.62^a	63.04 ± 5.45^a	36.96 ± 5.45^a	0.73 ± 0.28^a	69.56 ± 3.76^a	30.44 ± 3.76^a	0.51 ± 0.18^a	6.80 ± 0.50^a	0.008 ± 0.003^a
		$60 < MP \leq 90$	3.41 ± 0.35^b	4.83 ± 1.31^b	65.56 ± 3.88^{ab}	34.44 ± 3.88^{ab}	1.68 ± 0.55^b	69.03 ± 3.78^a	30.97 ± 3.78^a	1.16 ± 0.38^b	6.93 ± 0.43^a	0.018 ± 0.006^b
		$90 < MP \leq 120$	4.46 ± 0.13^c	8.24 ± 0.94^c	68.04 ± 4.39^b	31.96 ± 4.39^b	2.64 ± 0.53^c	69.99 ± 2.69^a	30.01 ± 1.69^a	1.85 ± 0.31^c	6.67 ± 0.49^a	0.028 ± 0.006^c
		$120 < MP \leq 140$	4.49 ± 0.04^c	10.31 ± 0.65^d	66.56 ± 3.42^b	33.44 ± 3.42^b	3.46 ± 0.52^d	69.92 ± 3.28^a	30.08 ± 3.28^a	2.42 ± 0.43^d	6.89 ± 0.44^a	0.037 ± 0.006^d
	Mean±sd	140	3.46 ± 1.01	5.76 ± 3.33	65.48 ± 4.74	34.52 ± 4.74	1.95 ± 1.13	69.56 ± 3.45	30.44 ± 3.45	1.36 ± 0.80	6.83 ± 0.47	0.021 ± 0.012
	F2	$30 < MP \leq 60$	2.20 ± 0.05^a	2.02 ± 0.76^a	63.72 ± 4.05^a	36.28 ± 4.05^a	0.74 ± 0.29^a	71.04 ± 4.34^a	28.96 ± 4.34^a	0.52 ± 0.21^a	6.79 ± 0.47^a	0.008 ± 0.003^a
		$60 < MP \leq 90$	3.38 ± 0.36^b	4.75 ± 1.07^b	67.65 ± 4.58^b	32.35 ± 4.58^b	1.53 ± 0.38^b	69.95 ± 3.85^a	30.05 ± 3.85^a	1.08 ± 0.29^b	6.86 ± 0.45^a	0.016 ± 0.004^b
		$90 < MP \leq 120$	4.42 ± 0.18^c	8.22 ± 1.08^d	68.39 ± 4.12^b	31.61 ± 4.12^b	2.59 ± 0.45^c	69.85 ± 3.74^a	30.15 ± 3.74^a	1.52 ± 0.40^d	6.89 ± 0.48^a	0.027 ± 0.005^c
		$120 < MP \leq 136$	4.49 ± 0.06^c	9.89 ± 0.60^d	65.15 ± 5.12^{ab}	34.85 ± 5.12^{ab}	3.43 ± 0.41^d	71.36 ± 3.34^a	28.64 ± 3.34^a	2.45 ± 0.36^d	6.68 ± 0.48^a	0.036 ± 0.005^d
	Mean±sd	136	3.47 ± 0.98	5.62 ± 3.10	66.27 ± 4.76	33.73 ± 4.76	1.86 ± 1.04	70.47 ± 3.89	29.53 ± 3.89	1.31 ± 0.74	6.82 ± 0.47	0.020 ± 0.011
	F3	$30 < MP \leq 60$	2.14 ± 0.40^a	1.83 ± 0.69^a	66.48 ± 3.91^a	33.52 ± 3.91^a	0.60 ± 0.20^a	71.29 ± 3.95^a	28.71 ± 3.95^a	0.43 ± 0.15^a	6.71 ± 0.47^a	0.007 ± 0.002^a
		$60 < MP \leq 90$	3.33 ± 0.34^b	4.31 ± 1.12^b	65.82 ± 4.11^a	34.18 ± 4.11^a	1.46 ± 0.39^b	70.23 ± 3.97^a	29.77 ± 3.97^a	1.02 ± 0.27^b	6.63 ± 0.56^a	0.015 ± 0.005^b
		$90 < MP \leq 120$	4.38 ± 0.21^c	7.86 ± 0.93^c	67.19 ± 4.84^a	32.81 ± 4.84^a	2.58 ± 0.50^c	70.45 ± 3.61^a	29.55 ± 3.61^a	1.82 ± 0.28^c	6.86 ± 0.47^a	0.027 ± 0.006^c
		$120 < MP \leq 137$	4.49 ± 0.08^c	10.25 ± 0.83^d	67.03 ± 3.44^a	32.97 ± 3.44^a	3.48 ± 0.41^d	71.73 ± 3.04^a	28.27 ± 3.04^a	2.43 ± 0.34^d	6.66 ± 0.44^a	0.037 ± 0.004^d
	Mean±sd	137	3.46 ± 0.99	5.50 ± 3.20	66.56 ± 4.13	33.44 ± 4.13	1.82 ± 1.08	70.83 ± 3.72	29.17 ± 3.72	1.29 ± 0.78	6.72 ± 0.49	0.020 ± 0.010
	Mean±sd	*140*	*3.47 ± 0.98*	*5.63 ± 3.20*	*66.10 ± 4.56*	*33.90 ± 4.56*	*1.88 ± 1.08*	*70.29 ± 3.72*	*29.71 ± 3.72*	*1.32 ± 0.77*	*6.79 ± 0.48*	*0.020 ± 0.010*
Total dataset	**Mean±sd**	**$30 < MP \leq 140$**	**3.46 ± 1.01**	**5.65 ± 3.23**	**65.98 ± 4.71**	**34.02 ± 4.71**	**1.90 ± 1.10**	**71.57 ± 4.30**	**28.43 ± 4.30**	**1.36 ± 0.80**	**6.74 ± 0.51**	**0.021 ± 0.012**

Different letters denote significant differences in pig manure characteristics and methane production among different masses at $p \leq 0.05$ based on Tukey's HSD post hoc test
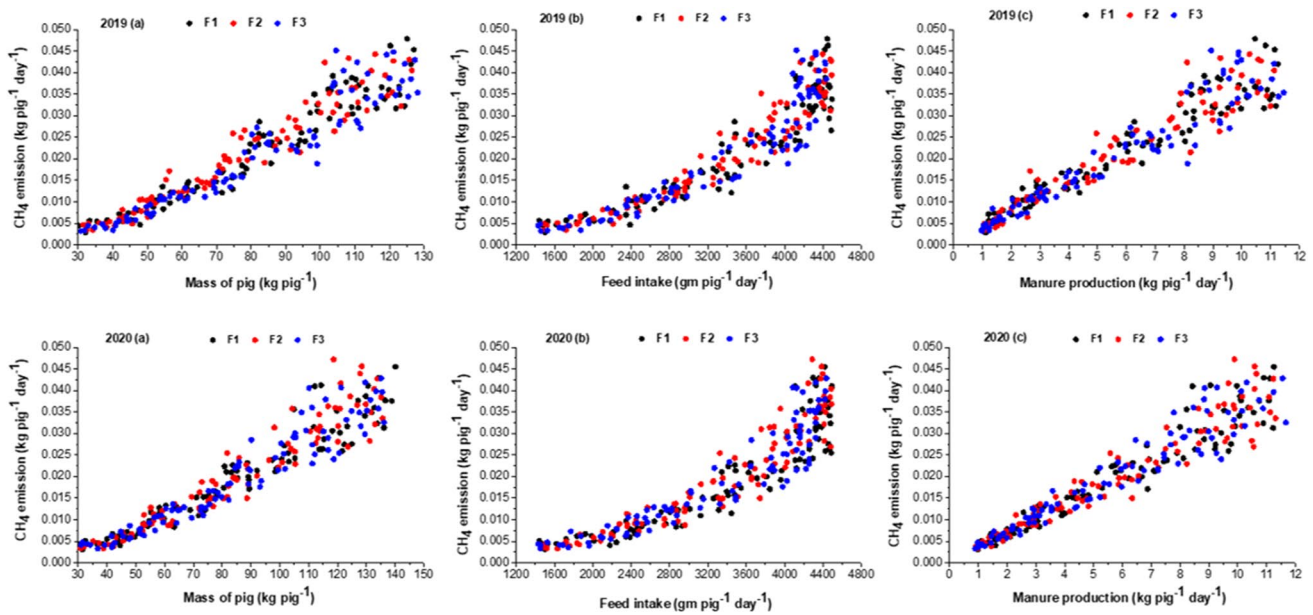
Fig. 3 Mass of pig, feed intake and manure production from pigs with three concentrated diets (F1, F2, and F3). 2019 **a** Mass of pig (kg pig^{-1}) vs. CH$_4$ emission rate (kg pig^{-1} day^{-1}) in 2019; 2019 **b** Feed intake (gm pig^{-1} day^{-1}) vs. CH$_4$ emission rate (kg pig^{-1} day^{-1}) in 2019; 2019. **c** Manure production (kg pig^{-1} day^{-1}) vs. CH$_4$ emis-sion rate (kg pig^{-1} day^{-1}) in 2019; 2020 (**a**): Mass of pig (kg pig^{-1}) vs. CH$_4$ emission rate (kg pig^{-1} day^{-1}) in 2020; 2020 **b** Feed intake (gm pig^{-1} day^{-1}) vs. CH$_4$ emission rate (kg pig^{-1} day^{-1}) in 2020; 2020 **c** Manure production (kg pig^{-1} day^{-1}) vs. CH$_4$ emission rate (kg pig^{-1} day^{-1}) in 2020

which was lower than the IPCC 2006 value specified for Oceania (11–13 kg pig^{-1} year^{-1}), Western Europe (6–21 kg pig^{-1} year^{-1}), and even North America (10–23 kg pig^{-1} year^{-1}) regions for market swine. One of the main reasons beyond this may be due to giving a more concentrated than forage diet in Korea. Several studies were conducted to find out the relation between diets and CH$_4$ emission rate (Hamilton et al. 1997; Kumar et al. 2014; Won et al. 2014). It is reported that feeding high forage diet leads to more CH$_4$ emission compared to a concentrated rich diet (Won et al. 2014). This is congruent with Beauchemin et al.'s (2008) findings for both diets, which also show that concentrated-based diets such as starch-rich grains produce less CH$_4$ than forage-based diets. It is found that concentrated diets reduce enteric CH$_4$ production by inhibiting the capacity of ruminal methanogens to take up hydrogen by reducing ruminal fluid pH and favoring the production of propionate over acetate (Van Kessel and Russell 1996; Pirondini et al. 2015). Pirondini et al. (2015) revealed that propionate production in the rumen decreases CH$_4$ production because propiogenesis uses metabolic hydrogen that would otherwise be available to produce CH$_4$. For instance, CH$_4$ production decreased up to 31% when 50% of dietary forage fed to dairy cows was replaced with a concentrated wheat grain diet (Moate et al. 2014). The daily minimum ruminal pH, which was linked to CH$_4$ generation, was responsible for the difference in CH$_4$ production (Moate et al. 2017). Moreover, it should be noted that the manure management system, environmental conditions, and storage time also influence the CH$_4$ emission (Wood et al. 2014).

Evaluating statistical algorithms for CH$_4$ emission modelling

The performance of two statistical models (multiple linear regression and polynomial regression) (F1, F2, F3, and FC) is shown in Table 4. Comparatively, the PR model performed better than the MLR in the training stage. However, the performance of the two models was almost similar in the testing stage. Table 4 indicated that PR-based statistical models yielded R^2 and RMSE values as 0.914 and 0.0035, which were the highest and lowest values, respectively in the training set as compared to the values obtained from MLR. Moreover, apart from the F1, F2, and F3 datasets, PR regression showed better results for the FC dataset (Table 4). The two models' overall performance in terms of R^2 is reasonably good, indicating that they explained more than 89% of the variations in the measured and predicted data. In general, the efficiency of all regression-based statistical models depends on the existence of linear relationships between explanatory and response variables. Due to their simplicity in nature, these models have been used widely to predict CH$_4$ emission (Petersen et al. 2016; Ngwabie et al. 2018; Hempel et al. 2020). An exponential equation was used for CH$_4$ emission modeling as a function of pigs' mass, where the models explained about 88% of the variations in the

Table 4 Performance metrics (R^2 and RMSE) of the models during testing and training period. The values in italics indicate the best results among the models

Model name	Dataset	Training		Testing	
		R^2	RMSE	R^2	RMSE
MLR	F1	0.897	0.0039	0.879	0.0040
	F2	0.913	0.0036	0.889	0.0039
	F3	0.901	0.0038	0.851	0.0045
	FC	0.901	0.0038	0.878	0.0041
PR (order-3)	F1	0.912	0.0036	0.872	0.0041
	F2	0.908	0.0037	0.876	0.0041
	F3	0.911	0.0036	0.863	0.0043
	FC	0.914	0.0035	0.890	0.0038
RR	*F1*	*0.918*	*0.0035*	*0.908*	*0.0035*
	F2	0.913	0.0036	0.902	0.0037
	F3	0.915	0.0035	0.905	0.0035
	FC	0.911	0.0036	0.897	0.0038
RFR	F1	0.980	0.0017	0.823	0.0048
	F2	0.992	0.0011	0.863	0.0043
	F3	0.986	0.0015	0.845	0.0046
	FC	0.982	0.0016	0.845	0.0046
ANN	F1	0.885	0.0042	0.852	0.0044
	F2	0.893	0.0040	0.865	0.0042
	F3	0.887	0.0041	0.819	0.0049
	FC	0.904	0.0037	0.865	0.0043

measured and predicted data (Ngwabie et al. 2018). Besides, the differences between the preceding and the current study were the selection of input variables and the algorithms used for CH_4 modelling. As shown in the scatter plots (Fig. 4), the predicted CH_4 emission rates have a very close distribution pattern with measured values with $R^2 = 0.892$, 0.894, 0.866, and 0.885 for F1, F2, F3, and FC, respectively during the testing set for MLR and the corresponding values of R^2 for PR model are 0.889, 0.879, 0.877, and 0.894.

Evaluating machine learning algorithms for CH_4 emission modelling

Determining the ML algorithm that best incorporates the mass of pigs and the quantity of feed intake into the CH_4 emission rate depends on the investigatory priority. The performance metrics of RR, RFR, and ANN are given in Table 4. The result showed that the RR model performed slightly better than RFR and ANN in the testing stage, whereas they were almost the same for the training stage. Table 4 indicated that the RR model yielded R^2 and RMSE values as 0.908 and 0.0035, which were the highest and lowest values, respectively in the testing set as compared to the values obtained from RFR and ANN. When the results of the ANN are compared to other machine learning models, ANN

performance was slightly lower. Moreover, the ANN model showed some better results using a large dataset; however, RFR was overestimated in the training stage and showed lower performance in the testing stage.

In contrast, the RR showed good performance ($R^2 \geq 0.90$ and RMSE ≤ 0.0038) even within the reduced dataset (Table 4). Several studies reported that in general, ANN and RFR models are more efficient and doing better performance compared to the RR method when there is a highly non-linear and complex relationship exist between output and inputs variables (Singh et al. 2003; Abdel-Rahman et al. 2013; Gholipoor et al. 2013; Khairunniza-Bejo et al. 2014; Mansourian et al. 2017; Qin et al. 2018; Basak et al. 2020). To better understand the distribution of data and the ML models' ability to predict the CH_4 emission rate, the predicted and measured data for testing datasets were presented and compared in the scatter plots (Fig. 4). As shown in these plots (Fig. 4), the predicted CH_4 emission rate has a very close distribution pattern with measured data, and both have almost the same pattern with R^2 values of 0.922, 0.904, 0.907, and 0.901 for F1, F2, F3, and FC dataset, respectively for the RR model. From this standpoint, RR would be preferred due to its simplicity and interpretability of the parameters and performed better, even using a small dataset compared to other ML algorithms in this study.

Model comparison and proposed model

All regression-based models performed better compared to the artificial neural network models in this study. The selected RR model could predict CH_4 emission rate for training and testing stages with a 2.50 and 6.20% increase in R^2 and a reduction of 11.25, and 17.98% in RMSE, respectively, compared with the ANN model. The superiority of the RR model to the other statistical and ML models is also well defined by considering standard deviation and correlation metric (Fig. 5). A graphical presentation of the actual and predicted values by the statistical and ML models over a graph (Fig. 5) could better understand these models' abilities. As shown in Fig. 4, there is somewhat a linear relationship between the input and output variables, which may be one of the main reasons for performance variation among those models.

Moreover, some other studies also showed that ANN models could not significantly improve the prediction accuracy compared to the statistical models due to the linear nature of variables (Özesmi et al. 2006; Craninx et al. 2008). The difference in performance among those models to predict the CH_4 emission rate showed the importance of choosing a proper model. According to the performance of those models, it can be concluded that the mass of pigs and the quantity of feed intake somewhat had a linear relationship with manure production and VS, which are closely

Fig. 4 Scatter plots of measured versus predicted CH_4 emission rate using F1, F2, F3, and FC datasets and MLR, PR, RR, RFR, and ANN models
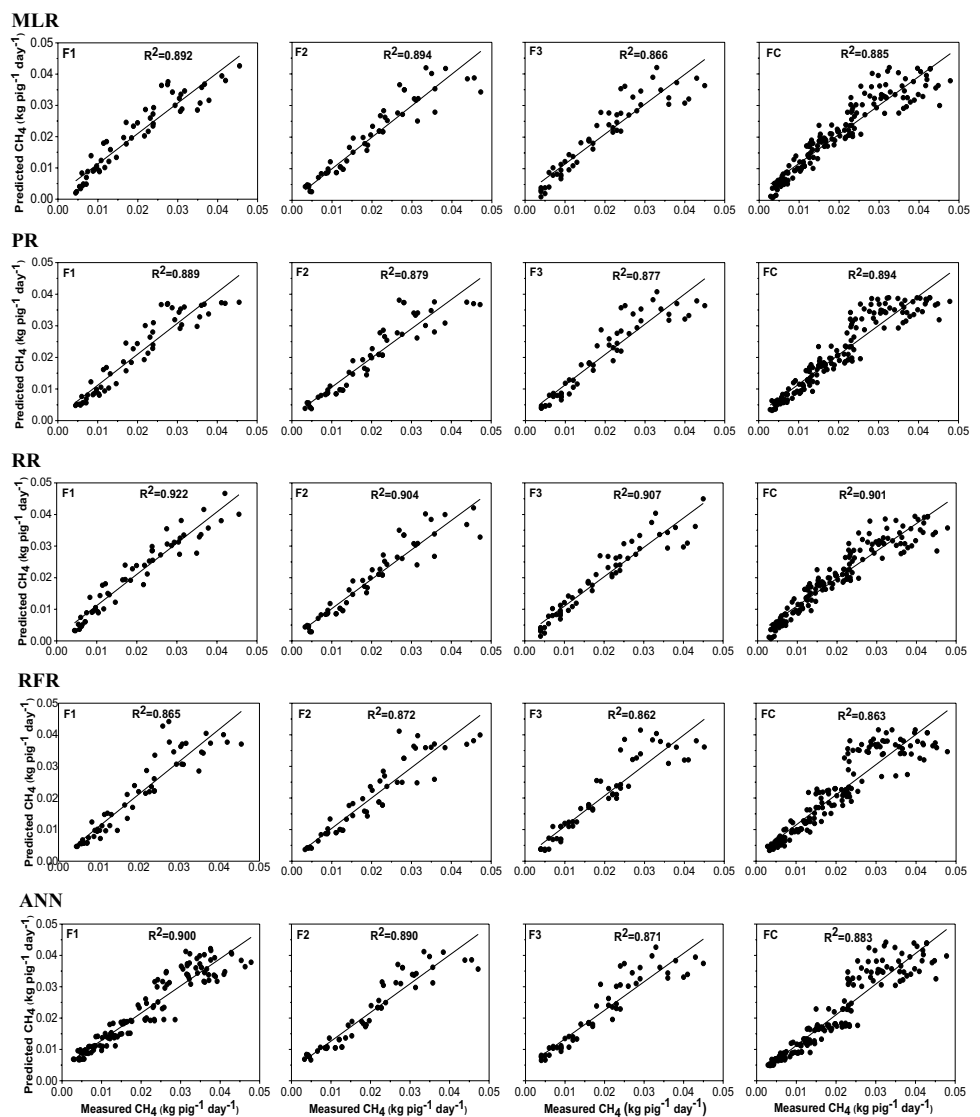
associated with the CH_4 emission rate. Therefore, in terms of developing the statistical and machine learning models in predicting CH_4 emission from livestock manure, the study recommends the use of regression-based algorithms to reveal more fruitful results.

Conclusion

Measurements were carried out in three experimental pig's barns with three different types of concentrated diets to characterize manure production. The quantity of manure produced per pig, moisture content, DM, ash, and VS contents increased with the mass and feed intake of pigs. Body mass ranged from 60 to 90 kg a pig produced around 3.35 kg of manure per day consisting of 66% moisture content and 34% DM. The manure's ash content was 28% DM (0.47 kg pig^{-1} day^{-1}), while the VS was 72% DM

(1.15 kg pig^{-1} day^{-1}). In the present study, the pigs' mass and the quantity of feed intake were used as explanatory variables to model the CH_4 production rate. Five statistical and ML algorithms were evaluated based on three statistical qualitative parameters for CH_4 emission modelling. The results showed that the regression-based models performed better than the ANN model. Moreover, the RR model was selected as the best model among those models in predicting CH_4 production. This priority for RR models may be because of the existing linear association between the mass of pigs and the quantity of feed intake with the CH_4 production rate. The RR model can explain more than 90% of the variations in all measured and predicted data in both the training and testing stages. With the ease of computing, simplicity, and interpretability of the parameters and better performance of RR models, it might be effective for CH_4 emission modelling. However, the above-described input parameters may not always be the same
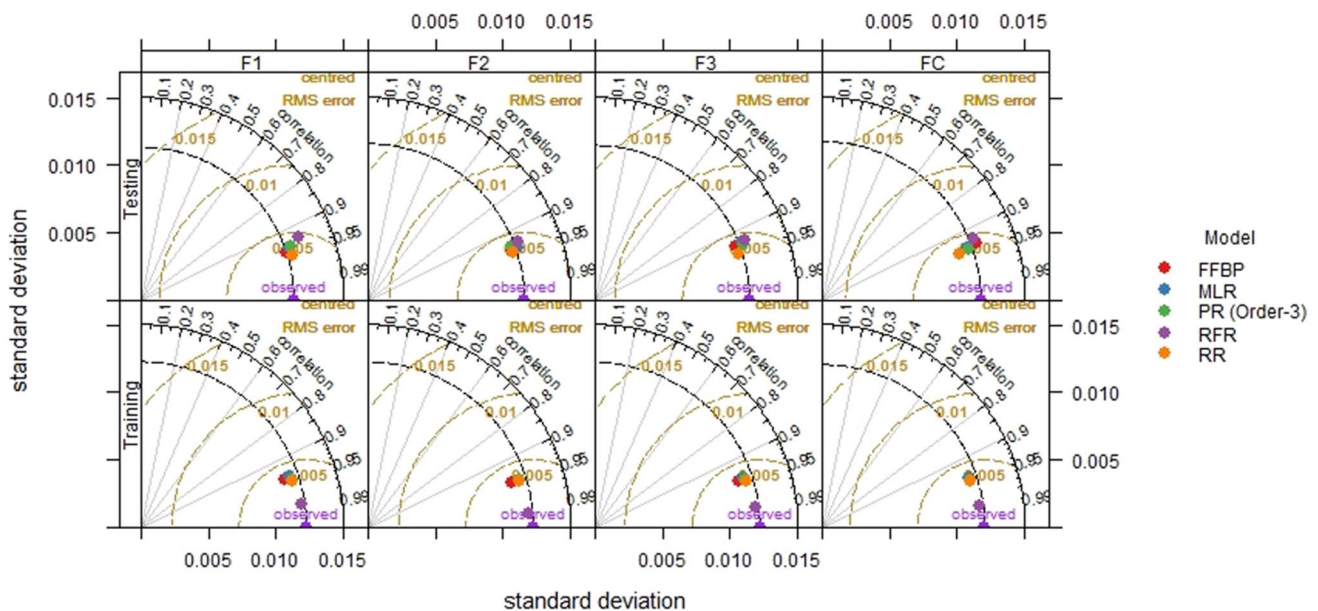
Fig. 5 Taylor diagram of training and testing results of statistical and machine learning models

when associated with CH_4 emission from pig manure. Additionally, trying to achieve high prediction efficiency of CH_4 emission modelling using the same attributes may lead to changes in the performance of the models. Therefore, research might be conducted to improve this model's prediction accuracy providing a wider range of diets and management conditions.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate The current research was conducted at the smart farm systems laboratory of Gyeongsang National University. The Institutional Animal Care and Use Committee (IACU) (GNU-150508-R0029) approved the experimental procedure and collection of data. All authors followed the ethical guidelines and safety procedures carefully during the experimental period.

Consent for publication All authors have read and agreed to the published version of the manuscript in the journal of Air Quality, Atmosphere, and Health.

Conflict of interest The authors declare no competing interests.

References

Abdel-Rahman EM, Ahmed FB, Ismail R (2013) Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. Int J Remote Sens 34(2):712–728. https://doi.org/10.1080/01431161.2012.713142

ANIR (2012). Australian national inventory report. Australian national greenhouse accounts: national inventory report. Canberra, ACT: Department of climate change and energy efficiency, Commonwealth of Australia

Basak JK, Qasim W, Okyere FG, Khan F, Lee YJ, Park J, Kim HT (2019) Regression analysis to estimate morphology parameters of pepper plant in a controlled greenhouse system. J Biosyst Eng 44:57–68. https://doi.org/10.1007/s42853-019-00014-0

Basak JK, Okyere FG, Arulmozhi E, Park J, Khan F, Kim HT (2020) Artificial neural networks and multiple linear regression as potential methods for modelling body surface temperature of pig. J Appl Anim Res 48(1):207–219. https://doi.org/10.1080/09712 119.2020.1761818

Beauchemin KA, Kreuzer M, O'Mara F, McAllister TA (2008) Nutritional management for enteric methane abatement: a review. Aust J Exp Agric 48:21–27. https://doi.org/10.1071/EA07199

Beninger PG, Boldina I (2014) Fine-scale spatial distribution of the temperate in faunal bivalve Tapes (=Ruditapes) philippinarum (Adams and Reeve) on fished and unfished intertidal mudflats. J Exp Mar Biol Ecol 457:128–134. https://doi.org/10.1016/j. jembe.2014.04.001

Biau G, Devroye L, Lugosi G (2008) Consistency of random forests and other averaging classifiers. J Mach Learn Res 9:2015–2033. https://www.jmlr.org/papers/volume9/biau08a/biau08a. pdf. Accessed 10 Dec 2020

Biau G, Scornet E (2016) A random forest guided tour. TEST 25:197–227. https://doi.org/10.1007/s11749-016-0481-7

Biau G (2012) Analysis of a random forests model. J Mach Learn Res 13:1063–1095. https://www.jmlr.org/papers/volume13/biau12a/ biau12a.pdf. Accessed 10 Dec 2020

Boilot P, Hines EL, Gardner JW, Pitt R, John S, Mitchell J, Morgan DW (2002) Classification of bacteria responsible for ENT and eye infections using the cyranose system. IEEE Sens J 2(3):247–253. https://doi.org/10.1109/JSEN.2002.800680

Carey N, Sigwart JD, Richards JG (2013) Economies of scaling: more evidence that allometry of metabolism is linked to activity, metabolic rate and habitat. J Exp Mar Biol Ecol 439:7–14. https://doi.org/10.1016/j.jembe.2012.10.013

Chastain JP, Camberato JJ, Albrecht JE, Adams J (1999) Swine manure production and nutrient content., South Carolina confined animal manure managers certification program. Chapter 3. Clemson, S.C: Clemson University. https://www.clemson.edu/ extension/camm/manuals/swine/sch3a_03.pd. Accessed 10 Dec 2020

Cortus E, Jacobson LD, Hetchler BP, Heber AJ, Bogan BW (2015) Methane and nitrous oxide analyzer comparison and emissions from dairy freestall barns with manure flushing and scraping. Atmos Environ 100:57–65. https://doi.org/10.1016/j.atmosenv. 2014.10.039

Cranford PJ, Ward JE, Shumway SE (2011). Bivalve filter feeding: variability and limits of the aquaculture biofilter. In SE Shumway (Ed.), Shellfish aquaculture and the environment. Hoboken: Wiley-Blackwell

Craninx M, Fievez V, Vlaeminck B, B. Baets B (2008) Artificial neural network models of the rumen fermentation pattern in dairy cattle. Comput Electron Agric 60(2):226-238. https://doi. org/10.1016/j.compag.2007.08.005

Darlington RB, Hayes, AF (2016) Regression Analysis and Linear Models: Concepts, Applications, and Implementation. Guilford Publications

Denil M, Matheson D, Freitas ND (2014) Narrowing the gap: random forests in theory and in practice. In Proceedings of the 31st International Conference on Machine Learning (ICML) 665–673

Dennehy C, Lawlor PG, Jiang Y, Gardiner GE, Xie S, Nghiem LD, Zhan X (2017) Greenhouse gas emissions from different pig manure management techniques: a critical analysis. Front Environ Sci Eng 11(3):1–16. https://doi.org/10.1007/ s11783-017-0942-6

Du Toit L, Niekerk W, Meissner HH (2013) Direct methane and nitrous oxide emissions of monogastric livestock in South Africa. S Afr J Anim Sci 43(3):362–375. https://doi.org/10.4314/sajas.v43i3.9

Ekine-Dzivenu CC, Mrode R, Oyieng E, Komwihangilo D, Lyatuu E, Msuta G, Ojango JMK, Okeyo AM (2020) Evaluating the impact of heat stress as measured by temperature-humidity index (THI) on test-day milk yield of small holder dairy cattle in a sub-Sahara African climate. Livest Sci 242:1–7. https://doi.org/10.1016/j. livsci.2020.104314

Environment-Canada (2012) National inventory report 1990–2010: greenhouse gas sources and sinks in Canada-executive summary. Government of Canada

FAO (2006) Livestock's long shadow. In Steinfeld H, Gerber P, Wassenaar T, Castel V, Rosales M, de Haan, C (Eds.) Environmental issues and options. Food and Agricultural Organization (FAO)

FAO (2011) World Livestock 2011–Livestock in Food Security. Food and Agricultural Organization (FAO), Rome, Italy

Feng X, Li Q, Zhu Y, Hou J, Jin L, Wang J (2015) Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. Atmos Environ 107:118–128. https://doi.org/10.1016/j.atmosenv.2015.02.030

Font-i-Furnols M, Terré M, Brun A, Vidal M, Bach A (2021) Prediction of tissue composition of live dairy calves and carcasses by computed tomography. Livest Sci 243:1–7. https://doi.org/10. 1016/j.livsci.2020.104371

Friedman J, Hastie T, Tibshirani R (2009) The elements of statistical learning. Springer Series in Statistics. Springer (NY), second edition. https://doi.org/10.1007/978-0-387-84858-7

Gholipoor M, Rohani A, Torani S (2013) Optimization of traits to increasing barley grain yield using an artificial neural network. Int J Plant Prod 7(1):1–18. https://doi.org/10.22069/IJPP.2012.918

Hamilton DW, Luce WG, Heald AD (1997) Production and characteristics of swine manure. Oklahoma state university extension fact No. F-1735

Hempel S, Adolphs J, Landwehr N, Willink D, Janke D, Amon T (2020) Supervised Machine Learning to Assess Methane Emissions of a Dairy Building with Natural Ventilation. Appl Sci 10(19):1–21. https://doi.org/10.3390/app10196938

Hirst AG (2012) Intra specific scaling of mass to length in pelagic animals: ontogenetic shape change and its implications. Limnol Oceanogr 57(5):1579–1590. https://doi.org/10.4319/lo.2012.57.5. 1579

Hoerl AE, Kennard RW (2000) Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 42(1):80–86. https:// doi.org/10.2307/1267351

Hunter JD (2007) Matplotlib: A 2D graphics environment. Comput Sci Eng 9(3):90–95. https://doi.org/10.1109/MCSE.2007.55

Hydrology (2000) ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. Artificial neural networks in hydrology. I: Preliminary concepts. J Hydrol Eng 5(2):115–123. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(115)

IPCC (2006) Emissions from livestock and manure management. In Dong M, Mangino J, McAllister TA, Hatfield, JL, Johnson DE, Lassey KR, de Lima MA, and Romanovskaya A. (Eds.), Guidelines for national greenhouse gas inventories. Intergovernmental Panel on Climate Change (IPCC)

Jaihuni M, Basak JK, Khan F, Okyere FG, Arulmozhi E, Bhujel A, Lee DH, Park J, Kim HT (2020) A Partially Amended Hybrid Bi-GRU-ARIMA Model (PAHM) for Predicting Solar Irradiance in Short and Very-Short Terms. Energies 13(2):1–20. https://doi. org/10.3390/en13020435

Ji ES, Park KH (2012) Methane and Nitrous Oxide Emissions from Livestock Agriculture in 16 Local Administrative Districts of Korea. Asian-Australas J Anim Sci 25(12):1768–1774. https:// doi.org/10.5713/ajas.2012.12418

Joharestani MZ, Cao C, Ni X, Bashir B, Talebiesfandarani S (2019) PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. Atmosphere 10(7):373. https://doi.org/10.3390/atmos10070373

Khairunniza-Bejo S, Mustaffha S, Ismail WIW (2014) Application of artificial neural network in predicting crop yield: a review. J Food Sci Eng 4:1–9

Kumar S, Choudhury PK, Carrod MD et al (2014) New aspects and strategies for methane mitigation from ruminants. Appl Microbiol Biotechnol 98(1):31–44. https://doi.org/10.1007/s00253-013-5365-0

Li C, Salas W, Zhang R, Krauter C, Rotz A, Mitloehner F (2012) Manure-DNDC: a biogeochemical process model for quantifying greenhouse gas and ammonia emissions from livestock manure systems. Nutr Cycl Agroecosyst 93:163–200. https://doi.org/10.1007/s10705-012-9507-z

Liang Y, Xin H, Wheeler EF et al (2005) Ammonia emissions from U.S. Laying hen houses in Iowa and Pennsylvania. T ASABE 48(5):1927–1941. https://doi.org/10.13031/2013.20002

Lu RD, Li YE, Shi F, Wan YF (2008) Effect of compost on the greenhouse gases emission from dairy manure. J Agro-Environ Sci 27(3):1235–1241

Mansourian S, Darbandi EI, Mohassel MHR, Rastgoo M, Kanouni H (2017) Comparison of artificial neural networks and logistic regression as potential methods for predicting weed populations on dry land chickpea and winter wheat fields of Kurdistan province. Iran Crop Prot 93:43–51. https://doi.org/10.1016/j.cropro.2016.11.015

McDonald GC (2009) Ridge regression. Wiley Interdisciplinary Reviews: Comput Stat 1(1):93–100. https://doi.org/10.1002/wics.14

McKinney W (2010) Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June-3 July 2010. https://doi.org/10.25080/Majora-92bf1922-012

Moate PJ, Williams SRO, Jacobs JL, Hannah MC, Beauchemin KA, Eckard RJ, Wales WJ (2017) Wheat is more potent than corn or barley for dietary mitigation of enteric methane emissions from dairy cows. J Dairy Sci 100:7139–7153. https://doi.org/10.3168/jds.2016-12482

Moate PJ, Williams SRO, Deighton MH, Wales WJ, Jacobs JL (2014) Supplementary feeding of wheat to cows fed harvested pasture increases milk production and reduces methane yield. In Proc. 6th Australasian Dairy Science Symposium, Hamilton, New Zealand. The Australasian Dairy Science Symposium Committee, Hamilton, New Zealand.

Ngwabie NM, Chungong BN, Yengong FL (2018) Characterisation of pig manure for methane emission modelling in Sub-Saharan Africa. Biosyst Eng 170:31–38. https://doi.org/10.1016/j.biosystemseng.2018.03.009

Ogejo JA (2019) Livestock and poultry environmental learning community. https://lpelc.org/manure-production-and-characteristics/. Accessed 23 December 2020

Olesen JE, Schelde K, Weiske A, Weisbjerg MR, Asman WAH, Djurhuus J (2006) Modelling greenhouse gas emissions from European conventional and organic dairy farms. Agric Ecosyst Environ 112(2–3):207–220. https://doi.org/10.1016/j.agee.2005.08.022

Ostertagová E (2012) Modelling using polynomial regression. Procedia Eng 48:500-506. https://doi.org/10.1016/j.proeng.2012.09.545

Özesmi SL, Tan CO, Özesmi U (2006) Methodological issues in building, training and testing artificial neural networks in ecological applications. Ecol Modell 195(1):83–93. https://doi.org/10.1016/j.ecolmodel.2005.11.012

Park KH, Andrew GT, Michele M, Karen C, Claudia W (2006) Greenhouse gas emissions from stored liquid swine manure in a cold climate. Atmos Environ 40(4):618–627. https://doi.org/10.1016/j.atmosenv.2005.09.075

Petersen SO, Olsen AB, Elsgaard L, Triolo JM, Sommer SG (2016) Estimation of Methane Emissions from Slurry Pits below Pig and Cattle Confinements. PLoS ONE 11(8):1–16. https://doi.org/10.1371/journal.pone.0160968

Pirondini M, Colombini S, Mele M, Malagutti L, Rapetti L, Galassi G, Crovetto GM (2015) Effect of dietary starch concentration and fish oil supplementation on milk yield and composition, diet digestibility, and methane emissions in lactating dairy cows. J Dairy Sci 98:357–372. https://doi.org/10.3168/jds.2014-8092

Qin Z, Myers DB, Ransom CJ et al (2018) Application of machine learning methodologies for predicting corn economic optimal nitrogen rate. J Agron 110(6):2596–2607. https://doi.org/10.2134/agronj2018.03.0222

Qu G, Feddes JJR, Armstrong WW, Coleman RN, Leonard JJ (2001) Measuring odor concentration with an electronic nose. T ASABE 44(6):1807–1812. https://doi.org/10.13031/2013.7018

Ransom CJ, Kitchen NR, Camberato JJ et al (2019) Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations. Comput Electron Agr 164:1–15. https://doi.org/10.1016/j.compag.2019.104872

Riaño B, García-González MC (2015) Greenhouse gas emissions of an on-farm swine manure treatment plant–comparison with conventional storage in anaerobic tanks. J Clean Prod 103:542–548. https://doi.org/10.1016/j.jclepro.2014.07.007

Rybarczyk Y, Zalakeviciute R (2018) Machine learning approaches for outdoor air quality modelling: a systematic review. Appl Sci 8(12):2570. https://doi.org/10.3390/app8122570

Seuront L (2010) Fractals and multifractals in ecology and aquatic science. Boca Raton: CRC Press, Taylor & Francis, Boca Raton. https://doi.org/10.1201/9781420004243

Shahriar SA, Kayes I, Hasan K, Salam MA, Chowdhury S (2020) Applicability of machine learning in modeling of atmospheric particle pollution in Bangladesh. Air Qual Atmos Health 13:1247–1256. https://doi.org/10.1007/s11869-020-00878-8

Shin HW, Llobet E, Gardner JW, Hines EL, Dow CS (2000) Classification of the strain and growth phase of cyanobacteria in potable water using an electronic nose system. IEE Proceedings - Science, Measurement and Technology 147(4):158–164. https://doi.org/10.1049/ip-smt:20000422

Shin J, Hong SG, Kim SC, Yang JE, Lee S, Li F (2016) Estimation of potential methane production through the mass balance equations from agricultural biomass in Korea. Appl Biol Chem 59:765–773. https://doi.org/10.1007/s13765-016-0224-1

Shin J, Hong SG, Kim SC, Yang JE, Lee SR, Li F (2017) Estimation of potential methane production through the mass balance equations from agricultural biomass in Korea. Appl Biol Chem 59(5):765–773. https://doi.org/10.1007/s13765-016-0224-1

Sinesio F, Di Natale C, Quaglia GB, Bucarelli FM, Moneta E, Macagnano A, Paolesse R, D'Amico A (2000) Use of electronic nose and trained sensory panel in the evaluation of tomato quality. J Sci Food Agric 80(1):63–71. https://doi.org/10.1002/(SICI)1097-0010(20000101)80:1%3c63::AID-JSFA479%3e3.0.CO;2-8

Singh T, Kanchan R, Verma A, Singh S (2003) An intelligent approach for prediction of triaxial properties using unconfined uniaxial strength. Miner Eng 5(2):12–16. https://doi.org/10.1007/s00521-012-1221-x

Steinfeld H, Gerber P, Wassenaar T, Castel V, Rosales M, de Haan C (2006) Livestock's Long Shadow: environmental issues and options. Food and Agricultural Organization, UN, Rome

Tabachnick B, Fidell LS (2001) Using multivariate statistics, 4th edition, Needham Heights, MA: Allyn & Bacon.

Telliard WA (2001) METHOD 1684: Total, fixed, and volatile solids in water, solids, and biosolids. U.S. Environmental Protection Agency

Tibshirani R (1996) Regression shrinkage and selection via the Lasso. J R Stat Soc Ser B 58(1):267–288

Tran TTK, Lee T, Shin J, Kim J, Kamruzzaman M (2020) deep learning-based maximum temperature forecasting assisted with meta-learning for hyperparameter optimization. Atmosphere 11(5):1–21. https://doi.org/10.3390/atmos11050487

US EPA (2020) Understanding Global Warming Potentials. https://www.epa.gov/ghgemissions/. Accessed 26 September 2020

Van Der Walt VDS, Colbert SC, Varoquaux G (2011) The NumPy array: a structure for efficient numerical computation. Comput Sci Eng 13(2):22–30. https://doi.org/10.1109/MCSE.2011.37

Van Kessel JAS, Russell JB (1996) The effect of pH on ruminal methanogenesis. FEMS Microbiol Ecol 20:205–210. https://doi.org/10.1016/0168-6496(96)00030-X

Wang J, Duan C, Ji Y, Sun Y (2010) Methane emissions during storage of different treatments from cattle manure in Tianjin. J Environ Sci 22(10):1564–1569. https://doi.org/10.1016/S1001-0742(09)60290-4

Wieringen WN (2020) Lecture notes on ridge regression. Version 0.31, July 17, 2020. Department of Epidemiology and Data Science, Amsterdam Public Health research institute, Amsterdam UMC, location VUmc. https://arxiv.org/pdf/1509.09169.pdf. Accessed 5 December 2020

Won SG, Cho WS, Lee JE, Park KH, Ra CS (2014) Data build-up for the construction of korean specific greenhouse gas emission inventory in livestock categories. Asian-Australas J Anim Sci 27(3):439–446. https://doi.org/10.5713/ajas.2013.13401

Wood JD, VanderZaag AC, Wagner-Riddle C, Smith EL, Gordon RJ (2014) Gas emissions from liquid dairy manure: complete versus partial storage emptying. Nutr Cycl Agroecosyst 99:95–105. https://doi.org/10.1007/s10705-014-9620-2

Zhang H, Balaban MO, Principe JC (2003) Improving pattern recognition of electronic nose data with time-delay neural networks. Sens Actuators B Chem 96(1–2):385–389. https://doi.org/10.1016/S0925-4005(03)00574-4