water

MDPI

Article

Modelling Methane and Nitrous Oxide Emissions from Rice Paddy Wetlands in India Using Artificial Neural Networks (ANNs)

Tabassum-Abbasi [1], Tasneem Abbasi [2,*], Chirchom Luithui [2] and Shahid Abbas Abbasi [2]

[1] Department of Health, Safety, Environment and Civil Engineering, University of Petroleum and Energy Studies, Dehradun 248 007, India; tab.abbasi@gmail.com

[2] Centre for Pollution Control and Environmental Engineering Pondicherry University, Chinakalapet Puducherry 605 014, India; chirluithui@gmail.com (C.L.); abbasi.cpee@gmail.com (S.A.A.)

* Correspondence: tasneem.abbasi@gmail.com

Received: 3 September 2019; Accepted: 13 October 2019; Published: 18 October 2019

Abstract: Paddy fields, which are shallow man-made wetlands, are estimated to be responsible for ~11% of the total methane emissions attributed to anthropogenic sources. The role of water use in driving these emissions, and the apportioning of the emissions to individual countries engaged in paddy cultivation, are aspects that have been mired in controversy and disagreement. This is largely due to the fact that methane (CH_4) emissions not only change with the cultivar type but also regions, climate, soil type, soil conditions, manner of irrigation, type and quantity of fertilizer added—to name a few. The factors which can influence these aspects also encompass a wide range, and have origins in causes which can be physical, chemical, biological, and combinations of these. Exceedingly complex feedback mechanisms, exerting different magnitudes and types of influences on CH_4 emissions under different conditions, are operative. Similar is the case of nitrous oxide (N_2O); indeed, the present level of understanding of the factors which influence the quantum of its emission is still more patchy. This makes it difficult to even understand precisely the role of the myriad factors, less so model them. The challenge is made even more daunting by the fact that accurate and precise data on most of these aspects is lacking. This makes it nearly impossible to develop analytical models linking causes with effects vis a vis CH_4 and N_2O emissions from paddy fields. For situations like this the bioinspired artificial intelligence technique of artificial neural network (ANN), which can model a phenomenon on the basis of past data and without the explicit understanding of the mechanism phenomena, may prove useful. However, no such model for CH_4 or N_2O has been developed so far. Hence the present work was undertaken. It describes ANN-based models developed by us to predict CH_4 and N_2O emissions using soil characteristics, fertilizer inputs, and rice cultivar yield as inputs. Upon testing the predictive ability of the models with sets of data not used in model development, it was seen that there was excellent agreement between model forecasts and experimental findings, leading to correlations coefficients of 0.991 and 0.96, and root mean square error (RMSE) of 11.17 and 261.3, respectively, for CH_4 and N_2O emissions. Thus, the models can be used to estimate CH_4 and N_2O emissions from all those continuously flooded paddy wetlands for which data on total organic carbon, soil electrical conductivity, applied nitrogen, phosphorous and potassium, NPK, and grain yield is available.

Keywords: *Oryza sativa*; rice; paddy; greenhouse gas emissions; methane; nitrous oxide; artificial neural networks

1. Introduction

1.1. Rice Cultivation as a Major Source of Greenhouse Gas (GHG) Emissions

Rice is produced and consumed in quantities that far exceed the production-consumption of any other food grain in Asia [1]. It is also the third most consumed grain in the world [2]. Moreover, rice production has been estimated to consume more water, and cause more greenhouse gas (GHG) emissions, per unit mass of rice yield, than any other food grain [3]. About 500 million tonnes of rice is produced in the world every year for which 11% of the world's arable land area goes under paddy cultivation [4]. The situation in the paddy fields, which are usually continuously flooded, are conducive to anoxic conditions developing near the water-soil interface. This, in turn, may lead to anaerobic digestion of organic carbon present in the soil, generating an approximately 2:1 volume/volume mixture of methane (CH_4) and carbon dioxide (CO_2). The former has 34 and 86 times greater global warming potential (GWG), on 100-year and 10-year time horizons respectively, than CO_2 [5,6]. At certain stages of paddy cultivation nitrous oxide (N_2O) is generated which has about 300 times greater global warming potential than CO_2 [6]. The resulting contribution of rice cultivation to anthropogenic GHG emissions is immense. Of these the contribution of CH_4 alone is estimated to be about 11% of all the CH_4 emission attributed to anthropogenic sources [6]. Corresponding estimates of N_2O emissions are not available because much less information exists on cause-effect links vis a vis N_2O emissions.

1.2. The Great Paucity of Information

In recent decades, especially in the current century, increasing attention is being paid towards locating identifying the causes which enhance GHG emissions as also the ones that reduce it so that strategies can be worked out to minimize the influence of the former while maximizing the effect of the latter [7]. The estimates of CH_4 and N_2O from paddies and the apportioning of the responsibility for controlling the emissions on the regions cultivating paddy is also required for meaningful management of the GHGs. However this is fraught with intense controversy because the estimates of GHG emissions made so far have been based on broad brush-stroke extrapolation from select studies to an extremely large variety of paddy wetlands that differ from each other in terms of not just the cultivar type but also in terms of the agroclimatic regions, soil conditions, manner of irrigation, type and quantity of fertilizer added, etc. [8,9]. However, a large number of inter-connected factors are known to influence emission of GHGs from paddy fields and have origins in causes which can be physical, chemical, biological, and combinations of these. Exceedingly complex feedback mechanisms, exerting different magnitudes and types of influences on GHG emissions under different conditions, are operative. This makes it difficult to even understand precisely the role of the various factors that influence GHG emissions, less so model them and control them.

The challenge is made even more daunting by the fact that accurate and precise data on most of these aspects is lacking, with very little field data available on GHG emissions in the world's top rice producing states. China is the world's leading producer of rice but has contributed only 326 studies on CH_4 emission and 228 studies on N_2O emission from its rice fields so far [9]. India, which is very close behind China in its rice output, is way behind in awareness of GHG emissions from its rice fields—with only 112 and 33 studies on CH_4 and N_2O emissions, respectively, to date. The other three among the world's top five rice producing countries—Indonesia, Bangladesh, and Vietnam—have still fewer studies on this aspect [9]. In contrast several times larger volumes of data are available on the quantities of organic carbon and major nutrients nitrogen (N), phosphorous (P), and potassium (K) present in the paddy fields. In much lesser fraction of studies one or more of other variables such as soil redox potential, medium and micro-nutrients, water table, etc., have been studied. This crippling paucity of data makes it well-nigh to develop analytical models linking causes with effects vis a vis CH_4 and N_2O emissions from paddy fields.

1.3. The Potential of the Artificial Neural Networks (ANNs)

Given that the interaction between the factors and the quantification of their rates of interaction is very complex and does not lend itself to developing analytical models for GHG emissions from paddy wetlands, it becomes necessary to use alternative modelling techniques that can capture the underlying complex interactions of the factors quantitatively. For situations like this a bioinspired artificial intelligence technique, especially the technique of artificial neural networks (ANN), which can model a phenomenon on the basis of past data and without an a priori or explicit understanding of the mechanism of the phenomena, may prove useful. The work presented in this paper was based on this premise, and is the first of its kind. It has led to models with which CH_4 and N_2O emissions from continuously flooded paddy fields can be forecast if data on organic carbon, soil electrical conductivity (SEC), NPK nitrogen, phosphorous and potassium, content, and yield of the rice variety are available.

The ANN methodology is based on mimicking the way neurons play their role in the operation of the human brain. Just as neurons in the human brain process the available information, based on past training, to arrive at intelligent solutions, artificial (and virtual) neurons are 'trained' with computers on the basis of existing data so that they can then 'intelligently' process new data and generate solutions one seeks. The ANN technique, which has been made increasingly versatile and precise over the years, has been gaining popularity due to its ability to pick up the linear or non-linear relationships among data, identify complex patterns in datasets, and make adequately accurate estimations even in situations where the number of input variables are limited [10–12]. Further, ANNs can be used for the mapping of input to output data without knowing a priori relationship between that data. They can process noisy and incomplete information and still produce acceptably precise forecasts. The competence of the ANNs is further enhanced by their non-parametric nature. Due to this, the learning mechanism of the ANNs is independent of the structure of the data; hence the need for establishing prior distribution of data is obviated [12,13].

2. Past Attempts at Using ANNs to Model CH_4 and N_2O Emissions

2.1. The Prior Art

The first two reports on the use of ANNs in forecasting CH_4 emissions from paddy fields have come from Chen et al. [14] and Setiawan et al. [15]. Chen et al. [14] report two models, based on support vector regression (ε-SVR) and back propagation-artificial neural network (BP-ANN), developed on data generated from six monitoring points in paddy fields located near the South Dongting lake during July–October 2010. The input parameters required for the model were ground biomass, ambient temperature, soil temperature at 5 cm depth, soil pH, electrical conductivity, and moisture. The prediction accuracy of the measured flux (mg of CH_4 emitted per unit area per hour) by the model was assessed in terms of the mean relative error (MRE) and the root mean square error (RMSE). The results indicated that the predictive abilities of both the models were close to each other, with the ε-SVR model performing marginally better than the BP-ANN in terms of a higher correlation coefficient and lower MRE and RMSE.

Setiawan et al. [15] studied the effect of groundwater table, soil pH, and temperature on CH_4 emissions. Observations recorded from paddy fields maintained on the system of rice intensification (SRI) practice in the Nagrak Organic SRI Centre, Indonesia, during November 2009–February 2010, were taken for constructing the ANN model. The input parameters were volumetric water content (VWC), soil temperature, and pH. Their ANN model had reasonable predictive ability, with correlation coefficients of 0.72 and 0.70 for CH_4 and N_2O respectively. The same group of authors [16] then extended and refined their work, obtaining a better correlation between ANN predictions and actual data by substituting the groundwater level (GWL) for the VMC in the input parameters.

Hasanah et al. [17] developed ANN models to predict CH_4 and N_2O emissions from data acquired from controlled experiments on paddy cultivation at the Bogor Agricultural University, Indonesia, employing SRI techniques, during March–June 2010. The inputs for the models were air temperature,

soil moisture, and pH. The RMSE was found to be 0.4 and 0.19 and the coefficient of determination 0.96 and 0.82 for the CH_4 and N_2O fluxes respectively. Arif et al. [18] have used data on soil moisture, soil temperature, soil electrical conductivity, soil redox potential, and soil pH, from experiments conducted in Bogor, Indonesia, during March–June 2015, to develop ANN models to predict CH_4 and N_2O fluxes. The models were able to predict the fluxes with coefficients of determination of 0.84 and 0.76 respectively.

2.2. Limitations of the Prior Art

All the five studies recounted above, which represent the entire prior art on the modelling of greenhouse gas (GHG) emission from paddy fields using the ANN technique, have their utility. However, they are all confined to the limited data generated by the authors. The models developed by Setiawan et al. [15,16], Hasanah et al. [17], and Arif et al. [18] have been developed using data obtained from controlled field experiments conducted by the authors over a short duration. The studies by Setiawan et al. [15,16] and Hasanah et al. [17] are on the same Ciherang rice variety while Arif et al. [18] have not mentioned the cultivar employed in their study. The models developed by Chen et al. [14] appear to be based on observations made from typical rice fields, and could therefore be a better representative of actual field conditions and emissions thereof. However, it is not clear if only one rice variety was employed or several in their study. Thus, it is difficult to say as to what extent can the models of these authors be relied upon to forecast the GHG emissions when paddy cultivation is done under very different agroclimatic conditions from the ones existing during their controlled experiments [15–18], or from field studies from a single region as in the case of Chen et al. [14].

One more aspect of the existing models that may limit their applicability to only the rice varieties grown by the authors is that emissions during the different stages of growth of one or the other rice variety, instead of the total emissions over the entire sowing-to-yield life of the plant, have been used to construct the ANN model. This may not be the most accurate nor representative choice of models for estimating GHG emissions as the emissions from the rice plants vary considerably during different stages of growth, with different cultivars showing different patterns of emissions during different growth stages [19]. Some varieties show maximum emission during the vegetative phase while some others do so at the maturity stage. Additionally, the quantum of emissions changes noticeably with varying soil conditions and fertilizer treatments even for the same cultivar type [7].

In the present work we tried to overcome these limitations of the prior art.

3. Methodology

3.1. General

When the phenomenon to be modelled by the ANN technique is such that experimental observations concerning it are available, in adequate volume, it is customary to use the past data in model development as well as in the validation and the testing of the model. The advantage of this approach is that it enables the use of a large volume of data, generated by several authors at numerous locations, some of which may be far apart. It may be prohibitively expensive and impractical for a single group to experimentally generate data of similar depth and breadth.

The common strategy is to separate the past data randomly into three slots. About 70% of the data is put into the first slot which is used for training the ANN and developing the ANN-based model. The second and the third slots, each comprising of about 15% of the data, are then used, respectively, in validating and testing the model. Examples of this approach abound in the literature. For instance, Bhange et al. [20] have used 70%, 15%, and 15% of the past data (generated by others) for training, validating, and testing their ANN model. In a similar fashion Perez-Zarate et al. [21] have used 80% of the past data for ANN training, 10% for validation, and 10% of training. Chen et al. [12] have also used 70% of past data (generated by others) for model development and the remaining 30% for model validation without using any data of their own.

In the present work we went a step further in ascertaining the robustness of our CH_4 model by taking out, randomly, an extra set of five data (out of a total of 76 datasets available in prior art) for use in a second round of model testing before even feeding the remaining 71 datasets to start the modelling process. In the next step, on our asking, the software randomly split the 71 datasets into 49 (70%), 11 (15%), and 11 (15%) datasets for the purpose of training, validation, and testing of the model, respectively. We then used the five datasets (which had been pulled out at the outset), for a second round of testing. We proceeded in like manner in the case of the N_2O model.

Needless to say, the steps mentioned above to develop any ANN-based model are possible only when enough numbers of experimental observations are available (on the aspect being modelled) in the prior art. However, if one has to develop an ANN model of any phenomena not studied before, one will be required to first experimentally generate adequate volume of the required data. Thereafter the data will have to be separated in three sets for training, validation, and testing, as explained above. In the instant case adequate prior art was available and doing fresh experiments was not necessary.

3.2. Obtaining the Data Bank

Luithui [22] had collected all the data reported so far on Indian paddy fields. It was updated by Dhanuja et al. [9] and Dhanuja [7]. A very extensive search using the global databases SCOPUS and Web of Science, besides Google Scholar and the electronic repositories provided by India's Inflibnet and Shodhganga led to 112 reports which had covered CH_4 emission from Indian paddy fields. Twenty-eight of these reports cover N_2O emission in addition to CH_4. Only five while the rest 33 reports are confined to N_2O emissions. From these, 260 and 76 datasets on CH_4 and N_2O emissions, respectively were obtained.

3.3. Choice of Input Parameters

Unfortunately, there is no standard, or even similar, protocol used by different authors in their parameter selection. As a result, the variables studied by them differ widely in type as well as numbers. The aspects studied by most of the authors are total organic carbon (TOC), SEC, NPK input, and grain yield (tons per ha). Apart from the fact that only these are the parameters which have been studied by the largest number of authors, they also happen to represent, albeit tenuously, soil properties, nutrient inputs, and a characteristic of the cultivar (yield).

A total of 76 and 30 datasets of these parameters were available vis a vis methane and nitrous oxide emissions respectively. Tables 1 and 2 present the ranges of the values of the six parameters in the available datasets. These were used in the ANN training and validation process. CH_4 and N_2O emissions in their seasonal integrated flux (kg/ha) values were taken as the target data to train the ANNs.

Table 1. Range, average, and standard deviation of the parameters used for developing the artificial neural network (ANN) for modelling methane (CH_4) emissions.

Parameter	Range	Average
TOC (%)	0.218–1.61	0.627 ± 0.307
SEC (dS/m)	0.32–0.6	0.435 ± 0.072
Nitrogen (kg/ha)	40–150	82.211 ± 36.224
Phosphorous (kg/ha)	20–60	38.139 ± 13.715
Potassium (kg/ha)	20–120	37.237 ± 15.284
Yield (t/ha)	2.37–7.2	4.466 ± 1.123
SIF (kg/ha)	7.6–207.17	48.056 ± 44.907

Table 2. Range, average, and standard deviation of the parameters used for developing the ANN for modelling nitrous oxide (N_2O) emissions.

Parameter	Range	Average
TOC (%)	0.42–0.94	0.75 ± 0.2
SEC (dS/m)	0.37–0.5	0.44 ± 0.04
Nitrogen (kg/ha)	40–120	70.53 ± 34.78
Phosphorous (kg/ha)	20–60	29.91 ± 14.72
Potassium (kg/ha)	20–60	33.33 ± 15.16
Yield (t/ha)	2.584–6.9	4.31 ± 1.31
SIF (g/ha)	79.5–1894.6	1002.08 ± 476.28

3.4. Development of the ANN Model

The technique of ANN was first reported in its most rudimentary form in 1943 by McCulloch and Pitts [23]. Its development proceeded slowly for the next 43 years before it received a major boost with the development of the 'back-propagation' (BP) algorithm by Rumelhart et al. [24]. The 1980s was also the time there was a spurt in the development of faster and better computers. These twin happenings have caused tremendous intensification the ANN use in a very large number of areas.

The present work has utilized the multi-layer perceptron (MLP)-ANN protocol, supported by the Levenberg–Marquardt feed-forward BP algorithm. The modelling involved the following steps:

1. Data collection, identification, and pre-processing of the model input and the target data.
2. Creating a network in MATLAB R2017b software by selecting the appropriate training functions, adaption learning function, performance function, number of layers, number of neurons, and transfer function.
3. Training the model with a part of the input and output data.
4. Assessing the performance of the trained ANN model using the mean squared error (MSE).
5. Retraining the network in case the performance is not satisfactory, by adjusting the training parameters until a robust network is generated.
6. Validating the model performance with the remaining data not used for training, and assessing the accuracy of the model in simulating the actual output, in terms of various statistical measures.

The architecture of the MLP-ANN comprised of three distinct layers: an input layer, an output layer, and unspecified number of hidden layers between the input and the output layers. To train the ANN, data was introduced at the input layer. A hidden layer consisting of 20 neurons was the processing section of the model. It is here that the computation of the weighted sum of the input signals and combination with a bias was carried out. This formed the pre-activation signal for the hidden layer which was then transformed by the hidden layer activation function to form feed forward activation signals that leave the hidden layer. Activation functions were used to transform the activation level of the neurons, which is the weighted sum of the inputs, into a linear output signal. At the output layer, in a similar fashion, the hidden layer activation signals were modified by weights, biases, and an output layer activation function to form the network output. This output was compared with the desired target and the error between the two was calculated. The error associated with the output was propagated back through the model and the network parameters so that the weights and biases were adjusted accordingly. This manner of back propagation computation was continued for several iterations till a minimization of error occurred.

The BP algorithm was used to train feed forward neural networks or multilayer perceptions. It is a method to minimize the cost function by changing weights and biases in the network. To learn and make better predications, a number of epochs (training cycles) were executed where the error determined by the cost function was propagated backward by gradient descent until a sufficiently small error was achieved.

The back-propagation method trained the multilayer neural network by modifying its synaptic connection weights between layers to improve model performance based on the error correction learning function. As needed, the learning function was continuous and differentiable.

The ANN model architecture as used by us is shown in Figure 1.

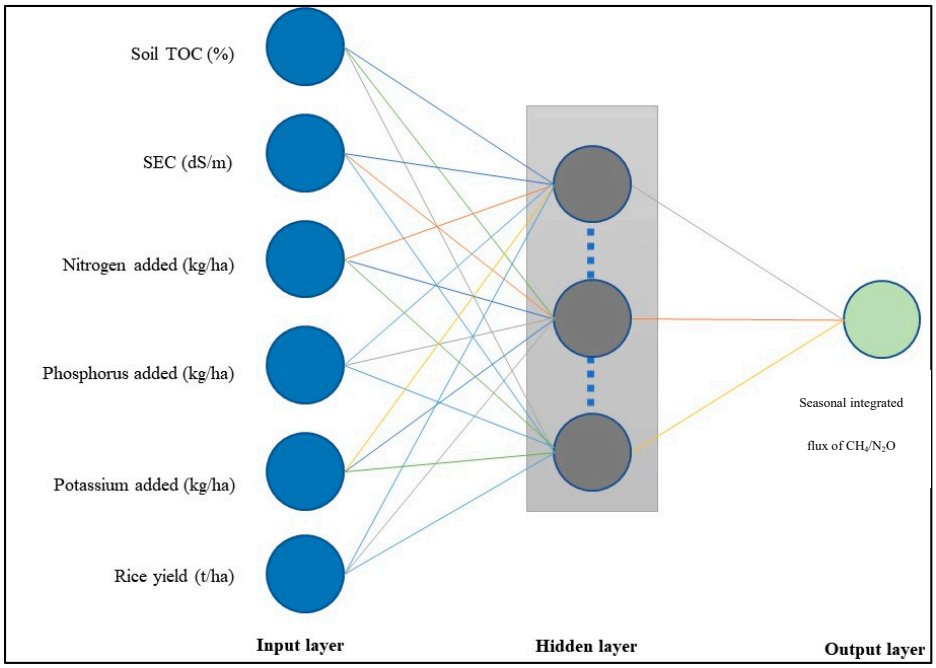

Figure 1. Schematic of the ANN model architecture.

As the system being modelled is quite complex, non-linear, and involves the interaction of many inter-related physical and chemical variables that exhibit spatial and temporal variation, we used a trial and error method to select the set of functions which could together give the most optimal network structure.

The feedforward network used by us constituted a single hidden layer with 20 neurons and 10 neurons respectively for CH_4 and N_2O emissions, and an output layer (Figure 2).

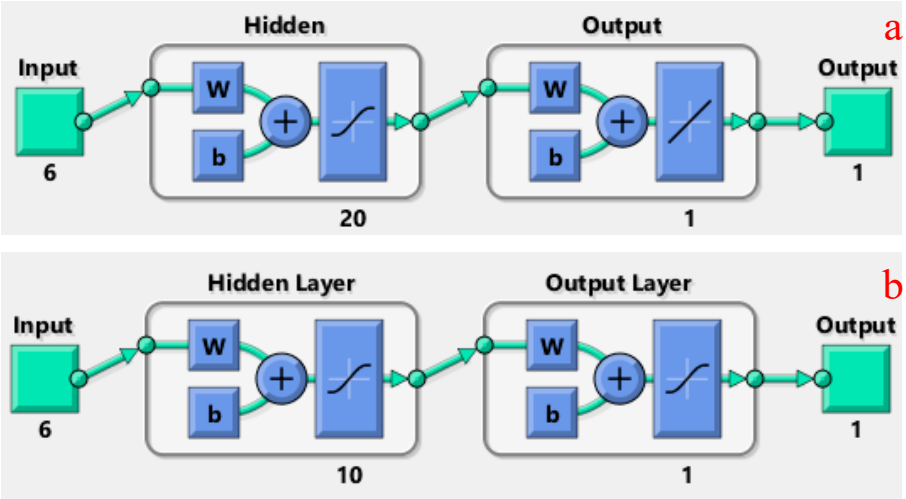

Figure 2. Structure of the artificial neural network used in this study (**a**) for CH_4 emissions (**b**) for N_2O emissions.

The hidden layer was activated by the tan-sigmoid transfer function tansig and output layer by the linear transfer functions purelin. The latter was chosen as it is best suited for solving fitting problems. The network was trained using trainlm, a function which updates weight and bias values according to Levenberg–Marquardt (LM) optimization. As trainlm calculates performance as the mean or sum of squared errors, the mean squared error (MSE) function was chosen as the performance goal function of the network.

3.4.1. Model Training Validation, and Testing

As explained in Section 3.1, of the 76 data sets that were obtained from prior art on methane emissions, five were randomly pulled out by us and kept aside for rechecking the model (after it had been validated and tested). The remaining 71 datasets were fed to the ANN software asking it to randomly partition the 71 datasets into sets of 49 (70%), 11 (15%), and 11 (15%) for the purposes of ANN training, model validation, and model testing, respectively. In other words, 49 datasets were used for the ANN training where the network was adjusted according to its error. Eleven datasets were used for validation to measure network generalization, and to halt training when generalization stopped improving as indicated by the increase in mean squared error of the validation samples. The remaining 11 datasets were used for testing. As these data sets had no bearing on the training of the network, they provided an independent measure of network performance during and after training. Thereafter the model performance was re-checked with the five datasets that had been pulled out before commencing the modelling process.

Similarly, for the modelling of nitrous oxide emissions, out of the 30 datasets that were obtained, 26 were used for development of the model and the remaining four were set aside for independent validation. Out of the 26 datasets used in modelling, randomly, 18 were randomly sectioned for training the model, four for validating the model, and the remaining four for testing the model.

3.4.2. Model Simulation

On retraining the network until the MSE was minimized and the correlation coefficients between the inputs and the targets were as close as possible to unity for all the datasets, five datasets from experimental analysis were fed into the trained network. The simulated output of the network was compared with that of the actual experimental values.

4. Results and Discussion

The robustness of the developed models was evaluated on the basis of the root mean square error (RMSE) as well as the correlation between the ANN-predicted and the measured fluxes. The results of the model training and testing are presented in Tables 3 and 4. The parity plot of the predicted versus measured fluxes indicating the correlation coefficient (R) between them is given in Figures 3 and 4. Histograms showing the distribution of the error (measured flux-predicted flux) is given in Figures 5 and 6.

Table 3. Performance of the ANN model for predicting CH_4 emissions, during the training and testing phases.

Stage of Development of the Model	Number of Datasets Taken	Root Mean Square Error (RMSE)	Correlation Between the Predicted and the Measured Flux
Training	49	17.74	0.922
Validation	11	26.65	0.898
Testing	11	16.80	0.854

Table 4. Performance of the ANN model for predicting N_2O emissions, during the training and testing phases.

Stage of Development of the Model	Number of Datasets Taken	Correlation Between the Predicted and the Measured Flux
Training	18	0.99
Validation	4	0.9998
Testing	4	0.998

As can be seen from Table 3 and Figure 3, the trained CH_4 emission model gave R values of 0.922 for the training phase, 0.898 for the validation phase, and 0.854 for the testing phase. The overall fit of the model in terms of the R value is 0.896. This indicates that the network has trained exceedingly well and has generated a model which achieves a close fit with the data.

In case of N_2O emissions also, the trained network is seen to perform remarkably well, despite very limited input data, yielding a strong correlation between the predicted and the measured fluxes. The correlation coefficients are close to unity (Table 4, Figure 4).

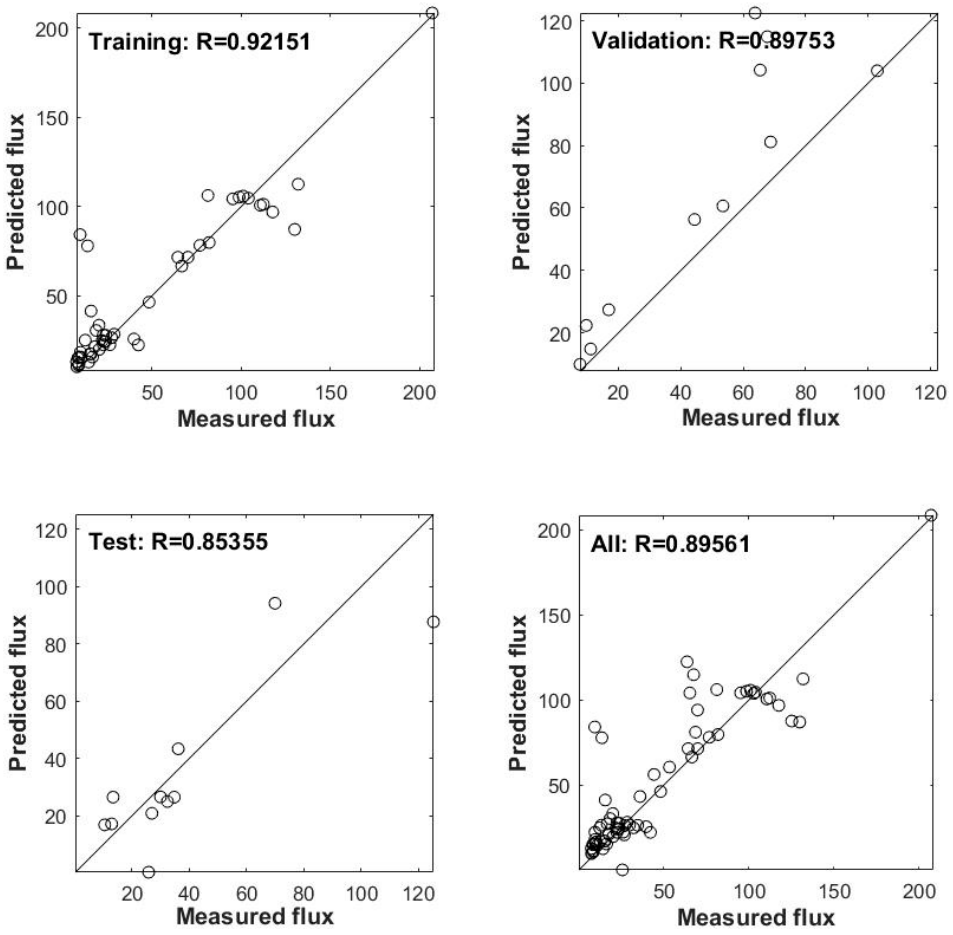

Figure 3. Network training and regression plots for the prediction of methane emissions.

Hence both the models can be used to estimate CH_4 and N_2O emissions that would have occurred in paddy cultivation during different stages of the past—and are likely to occur from new paddy fields—in which continuous flooding is practiced for irrigation. In recent decades increasing emphasis has been put on reducing water-use in paddy cultivation and encouraging intermittent irrigation. However, it has been claimed in a recent report [25], that even as intermittent irrigation helps in water saving and a reduction in CH_4 emissions in comparison to the classical way of continuous flooding

based paddy cultivation, the advantages are offset to a great extent by an increase in the N_2O emissions. The more recent findings of these authors [7,19] do not support the claims of Kritee et al. [25]. Evidently a lot more research is required before a broad consensus emerges on the role of water use in driving the net GHG emissions from paddy wetlands. The present work is expected to facilitate efforts in that direction.

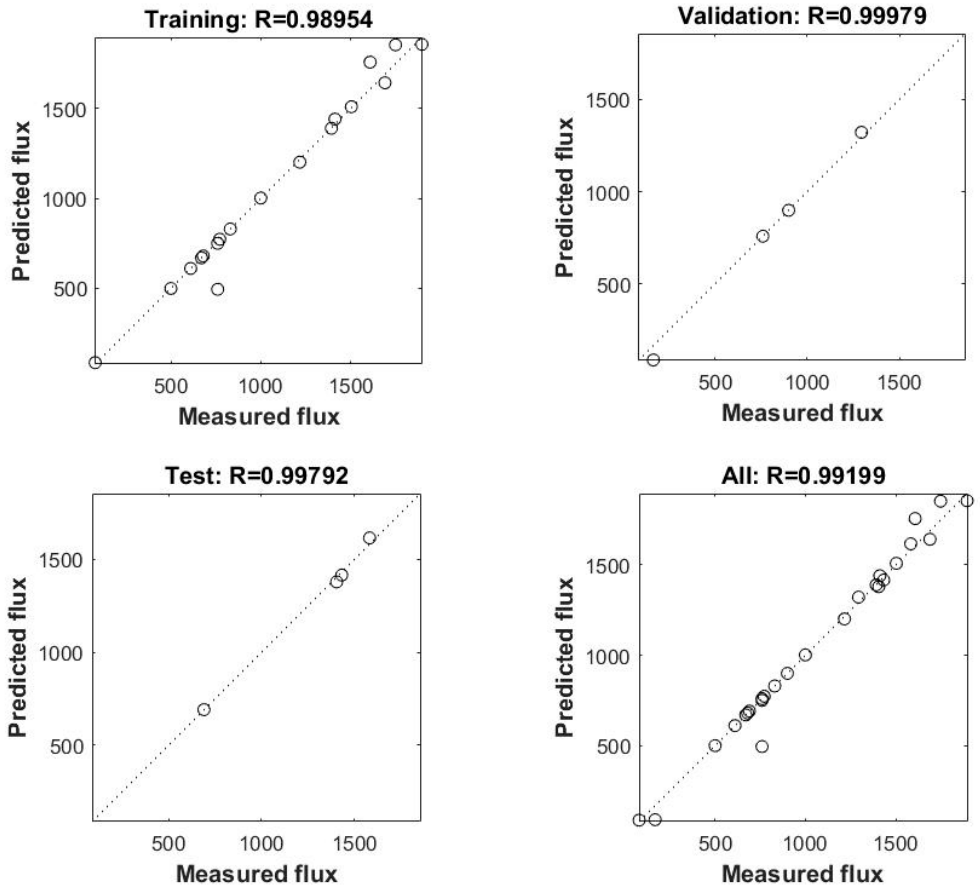

Figure 4. Network training and regression plots for the prediction of nitrous oxide emissions.

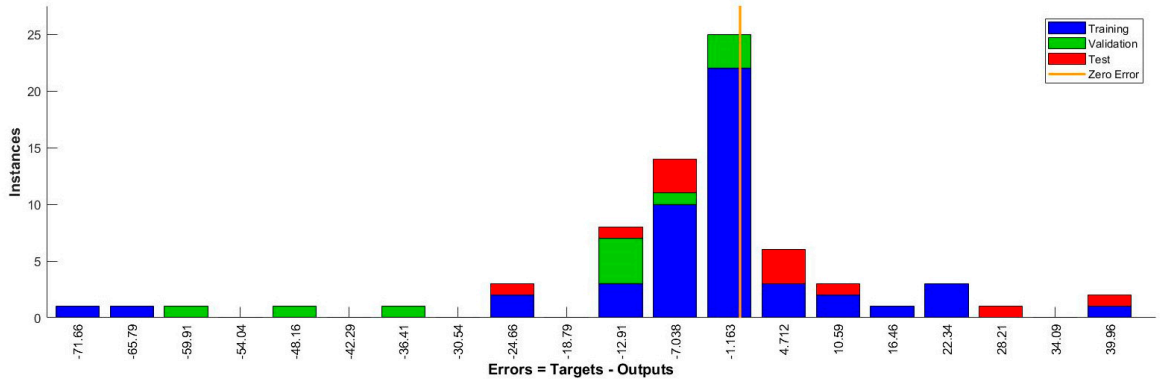

Figure 5. Error distribution during the development and testing of the ANN model for CH_4 emissions.

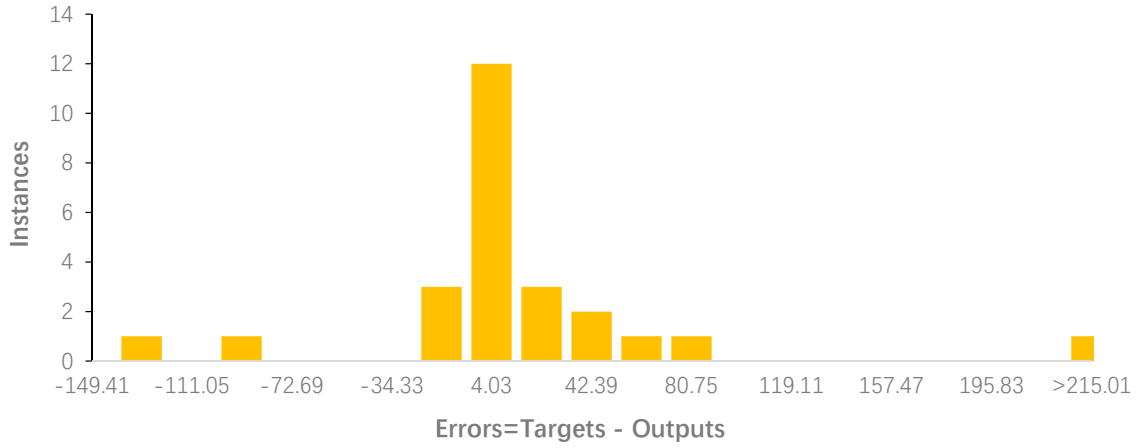

Figure 6. Error distribution during the development and testing of the ANN model for N_2O emissions.

From the histogram given in Figure 5 which depicts the distribution of the error (measured flux-predicted flux), it is seen that 73% of the predictions in the validation stage, and 82% of the predictions in the testing phase, fall in the error range of −25 to 25 (kg/ha). The largest error encountered in the testing phase was 40. When read together with the RMSE and the *R* values given in Table 2, these figures indicate that the network can be used to predict CH_4 emissions from paddy wetlands with a reasonable degree of accuracy.

From the error histogram of the nitrous oxide emissions model (Figure 6), it is seen that 62% of all the predictions lay in the error range of −25 to 25 (g/ha). It must be mentioned that while the methane emissions have been given in kg/ha, the nitrous oxide emissions have been measured in g/ha, and when seen in this context, the error range of −25 to 25 for nitrous oxide emissions is of smaller value compared to that of methane emission errors.

The performance of the ANN model was further assessed with the five datasets that were not included in the process of development of the model for methane emissions. The results are presented in Tables 5 and 6. The extent of deviation of the predicted flux from the measured flux in terms of percentage error ranged from −3.7% to 43%. The mean absolute error (MAE) came to 17.32%, and the RMSE was 11.17. The correlation coefficient was 0.991. The closeness of the predicted flux to the measured flux can also be seen from the parity plot (Figure 7) and the scatter plot (Figure 8).

Table 5. Fluxes predicted for five independent datasets that were not used for training the model.

Input Parameters						Measured Flux (kg/ha)	Predicted Flux (kg/ha)	Error (%)
TOC %	SEC	N (kg/ha)	P (kg/ha)	K (kg/ha)	Yield (t/ha)			
0.42	0.42	120	26.2	50	6.04	27	38.596	42.95
0.66	0.5	120	60	60	3.86	149.64	143.44	−4.14
0.66	0.5	90	60	60	5.02	187.2	198.17	5.86
0.86	0.5	60	30	30	3.11	27.14	26.126	−3.74
0.86	0.5	120	40	40	4.48	60.69	78.831	29.89

Table 6. Performance of the ANN model in predicting fluxes for the datasets given in Table 5.

Performance Measure	Value
Mean Absolute Error	17.32%
Root Mean Square Error	11.17
Correlation coefficient	0.991

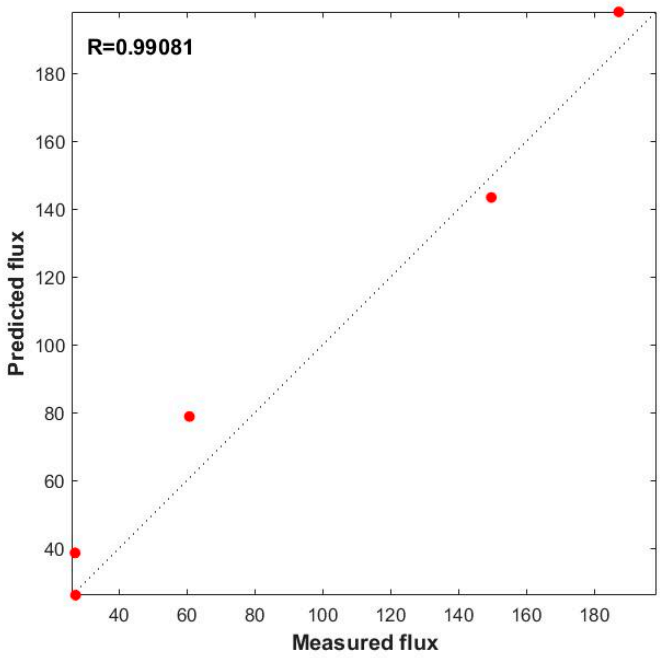

Figure 7. Performance of the ANN model in simulating the methane emissions from the five independent data sets not used for training and testing of the model.

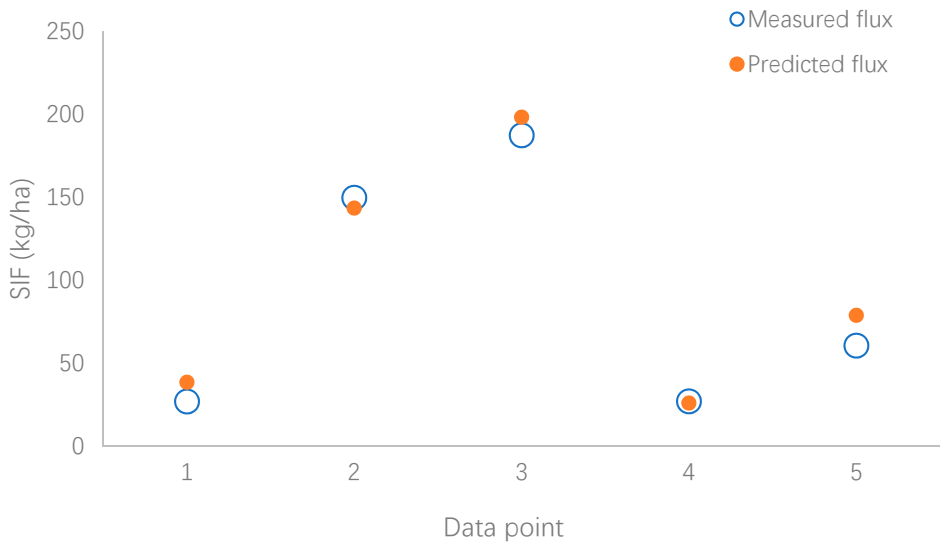

Figure 8. Scatter plot of the actual methane emissions versus the ANN predicted emission.

The model for nitrous oxide emissions was also validated with the four datasets that were not included in the process of development of the model. The results are presented in Tables 7 and 8. The extent of deviation of the predicted flux from the measured flux in terms of percentage error ranged from −7% to 50%. The mean absolute error (MAE) came to 21%, and the RMSE was 261.3. The correlation coefficient was 0.96. The closeness of the predicted flux to the measured flux can also be seen from the parity plot (Figure 9) and the scatter plot (Figure 10).

Table 7. Nitrous oxide fluxes predicted for four independent datasets that were not used for training the model.

Input Parameters						Measured Flux (kg/ha)	Predicted Flux (kg/ha)	Error (%)
TOC %	SEC	N (kg/ha)	P (kg/ha)	K (kg/ha)	Yield (t/ha)			
0.42	0.42	108	26.2	50	5.86	630	699.7641	11.07
0.66	0.5	90	60	60	4.32	570	529.3104	−7.134
0.78	0.5	80	40	40	4.66	660	955.6155	44.79
0.9	0.45	40	20	20	2.701	846.8	1270.128	49.99

Table 8. Performance of the ANN model in predicting nitrous oxide fluxes for the datasets given in Table 7.

Performance Measure	Value
Mean Absolute Error	21%
Root Mean Square Error	261.30
Correlation coefficient	0.9597

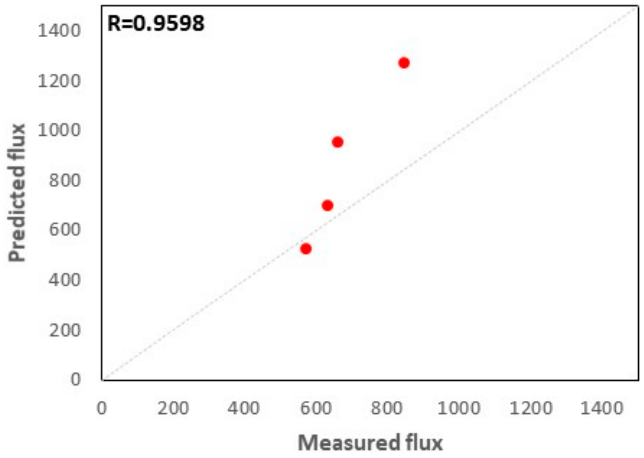

Figure 9. Performance of the ANN model in simulating the nitrous oxide emissions from the four independent datasets not used for training and testing of the model.

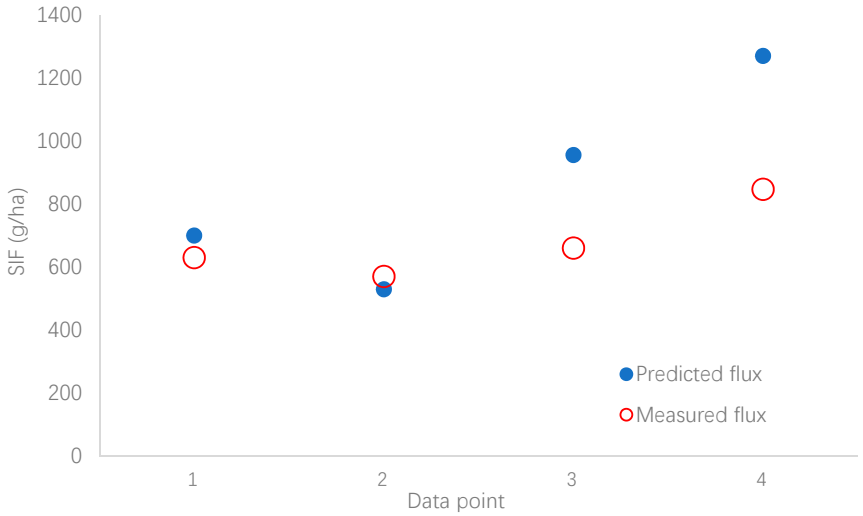

Figure 10. Scatter plot of the actual nitrous oxide emissions versus the ANN predicted emission.

5. Summary and Conclusions

The management of methane and nitrous oxide emissions from paddy wetlands is of great importance in the attempts to mitigate climate change. However, even as rice is the most widely cultivated and consumed grain in Asia, and third most in the world, there is a great paucity of information on the quantum of emissions of greenhouse gases from paddy fields. Consequently, there is sharp disagreement on the role of water in driving the net GHG emissions because continuous irrigation is implicated in high CH_4 generation while intermittent irrigation is believed to foster the release of N_2O. If models can be developed linking causes of GHG emissions with their quanta, it may become possible to use data on paddy cultivation to assess the likely GHG emissions even for situations when GHG measurements have not been made. However, so little is known about the causes, the relative extents of their contribution, and their mechanism that it is not possible to develop analytical models. In this backdrop the present work was undertaken in which the bioinspired artificial intelligence technique of artificial neural networks (ANN) was employed. A resort to this technique was taken because it has the capability to make forecasts of acceptable precision even when cause-effort relationships are not known, and the input data is noisy. All past data on factors influencing paddy cultivation in India, and associated GHG emissions was obtained. After identifying the five parameters which were covered in the largest number of reports, ANN models were developed linking those factors with fluxes of CH_4 or N_2O. Upon testing the predictive ability of the models vis a vis sets of data not used in model development, it was seen that there was excellent agreement between model forecasts and experimental findings, leading to correlations coefficients of 0.991 and 0.96, and RMSE of 11.17 and 261.3, respectively, for CH_4 and N_2O emissions.

Thus, the model can be used to estimate methane and nitrous oxide emissions from continuously flooded paddy wetlands for which data on total organic carbon, soil electrical conductivity, applied NPK, and grain yield is available. In this manner this model can contribute enormously in filling the huge data gaps prevailing in India on the quantum of methane and nitrous oxide emissions from continuously flooded paddy wetlands. Moreover, the same approach can be used to, equally gainfully, model the methane and nitrous oxide emissions in other regions of the world on the basis of data of that regions. Comparison of those information on CH_4 and N_2O emissions from intermittently flooded paddy cultivation can give insights into the relative impact of water use in enhancing or reducing either of the emissions.

Author Contributions: Conceptualization, T.-A., T.A., and S.A.A.; Methodology, T.-A., T.A., and S.A.A.; Software, T.-A., and T.A.; Validation, T.-A. and T.A.; Formal Analysis, T.-A., T.A., and S.A.A.; Investigation, C.L.; Resources, T.-A., T.A., and S.A.A.; Data Curation, C.L.; Writing-Original Draft Preparation, T.-A. and T.A.; Writing-Review & Editing, S.A.A.; Visualization, S.A.A.; Supervision, T.A. and S.A.A.; Project Administration, T.A.; Funding Acquisition, T.-A. and T.A..

Funding: Participation of S.A.A. was supported by the Emeritus Scientist grant (21(1034)/16/EMR-II) of Council of Scientific and Industrial Research (CSIR), New Delhi.

Acknowledgments: Support to T.-A., T.A., and C.L. was provided by their respective institutions. S.A.A. thanks the Council of Scientific and Industrial Research (CSIR), New Delhi, for the Emeritus Scientist grant (21(1034)/16/EMR-II)

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tsunoda, S.; Takahashi, N. *Biology of Rice*; Elsevier: Amsterdam, The Netherlands, 2012; Volume 7.
2. OECD/FAO Agricultural Outlook. OECD Agriculture Statistics, 2018. Available online: https://www.oecd-ilibrary.org/agriculture-and-food/data/oecd-agriculture-statistics_agr-data-en (accessed on 1 October 2019).
3. Wang, C.; Lai, D.Y.F.; Sardans, J.; Wang, W.; Zeng, C.; Penuelas, J. Factors Related with CH_4 and N_2O Emissions from a Paddy Field: Clues for Management implications. *PLoS ONE* **2017**, *12*, e0169254. [CrossRef] [PubMed]
4. IRRI (International Rice Research Institute). *Rice Almanac*, 4th ed.; Global Rice Science Parternship; IRRI: Las Bafios, Phillipines, 2013.
5. IPCC Working Group I. *Climate Change 2013-The Physical Science Basis: Summary for Policymakers*; Intergovernmental Panel on Climate Change: Geneva, Switzerland, 2013.
6. World Bank. *Greenhouse Gases from Reservoirs Caused by Biogeochemical Processes*; World Bank: Washington, DC, USA, 2017.
7. Dhanuja, C. Greenhouse Gas Emissions from Paddy Fields with Special Reference to the Effect of Vemricomposts and Synthetic Chemical Fertilizers. Ph.D. Thesis, Pondicherry University, Puducherry, India, 2019; pp. xiii+249.
8. Abbasi, S.A.; Dhanuja, C.; Abbasi, T. Emission of greenhouse gases from Indian wetlands: An overview. *Proc. Natl. Acad. Sci. India* **2019**. Communicated.
9. Dhanuja, C.; Saxena, D.K.; Abbasi, T.; Abbasi, S.A. Effect of application of vermicompost on methane emission and grain yield of *Chinna Ponni* paddy crop. *Paddy Water Environ.* **2019**, *17*, 797–802. [CrossRef]
10. Lu, F.; Chen, Z.; Liu, W.; Shao, H. Modeling chlorophyll-a concentrations using an artificial neural network for precisely eco-restoring lake basin. *Ecol. Eng.* **2016**, *95*, 422–429. [CrossRef]
11. Antonopoulos, V.Z.; Antonopoulos, A.V. Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. *Comput. Electron. Agric.* **2017**, *132*, 86–96. [CrossRef]
12. Chen, Z.; Ye, X.; Huang, P. Estimating Carbon Dioxide (CO_2) Emissions from Reservoirs Using Artificial Neural Networks. *Water* **2018**, *10*, 26. [CrossRef]
13. Ding, W.; Zhang, J.; Leung, Y. Prediction of air pollutant concentration based on sparse response back-propagation training feedforward neural networks. *Environ. Sci. Pollut. Res.* **2016**, *23*, 19481–19494. [CrossRef] [PubMed]
14. Chen, Q.; Jiang, W.; Chen, X.; Yuan, L.; Wang, W.; Pan, Y.; Wang, W.; Liu, X.; Liu, H. Prediciton of methane emissions of paddy field based on support vector regression model. *Environ. Sci.* **2013**, *34*, 2975–2982.
15. Setiawan, B.I.; Irmansyah, A.; Arif, C.; Watanabe, T.; Mirzoguchi, M.; Kato, H. Effects of groundwater level on CH_4 and N_2O emissions under SRI paddy management in Indonesia. *Taiwan Water Conserv.* **2013**, *61*, 135–146.
16. Setiawan, B.I.; Imansyah, A.; Arif, C.; Watanabe, T.; Mizoguchi, M.; Kato, H. SRI paddy growth and GHG emissions at various groundwater levels. *Irrig. Drain.* **2014**, *63*, 612–620. [CrossRef]
17. Hasanah, N.A.I.; Setiawan, B.I.; Mirzoguchi, M.; Sands, G.R.; Arif, C.; Widodo, S. Triangle graphs development for estimating methane and nitrous oxide gases emission from the system of rice intentsification (SRI). *J. Environ. Sci. Technol.* **2017**, *10*, 206–214. [CrossRef]
18. Arif, C.; Setiawan, B.I.; Hasanah, N.A.I.; Mizoguchi, M. Estimating Greenhouse Gas Emissions from Irrigated Paddy Fields in Indonesia under Various Water Managements. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *557*, 1–7. [CrossRef]
19. Dhanuja, C.; Abbasi, T.; Abbasi, S.A. Assessment of methane emission, and the factors which influence it, from three rice varieties commonly cultivated in the state of Puducherry. *Reg. Environ. Chang.* **2019**. Communicated.
20. Bhange, V.P.; Bhivgade, U.V.; Vaidya, A.N. Artificial Neural Network Modeling in Pretreatment of Garden Biomass for Lignocellulose Degradation. *Waste Biomass Valoriz.* **2019**, *10*, 1571–1583. [CrossRef]
21. Pérez-Zárate, D.; Santoyo, E.; Acevedo-Anicasio, A.; Díaz-González, L.; García-López, C. Evaluation of artificial neural networks for the prediction of deep reservoir temperatures using the gas-phase composition of geothermal fluids. *Comput. Geosci.* **2019**, *129*, 49–68. [CrossRef]

22. Luthui, C. Studies on Greenhouse Gas Emissions from Wetlands. Ph.D. Thesis, Pondicherry University, Puducherry, India, 2018; pp. xi+96.
23. McCulloch, W.S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–131. [CrossRef]
24. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
25. Kritee, K.; Nair, D.; Zavala-Araiza, D.; Proville, J.; Rudek, J.; Adhya, T.K.; Loecke, T.; Esteves, T.; Balireddygari, S.; Dava, O.; et al. High nitrous oxide fluxes from rice indicate the need to manage water for both long-and short-term climateimpacts. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 9720–9725. [CrossRef] [PubMed]