

Article

A Model to Forecast Methane Emissions from Tropical and Subtropical Reservoirs on the Basis of Artificial Neural Networks

Tabassum-Abbasi ¹, Tasneem Abbasi ^{2,*}, Chirchom Luithui ² and S. A. 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 & Environmental Engineering, Pondicherry University, Chinakalapet, Puducherry 605 014, India; chirluithui@gmail.com (C.L.); abbasi.cpee@gmail.com (S.A.A.)

* Correspondence: tasneemabbasi.cpe@pondiuni.edu.in

Received: 15 October 2019; Accepted: 30 December 2019; Published: 2 January 2020



Abstract: In view of the great paucity of information on the exact contributions of different causes which lead to different extents of emission of the greenhouse gas methane (CH₄) from reservoirs, it is tremendously challenging to develop statistical or analytical models for forecasting such emissions. Artificial neural networks (ANNs) have the ability to discern linear or non-linear relationships despite very limited data inputs and can recognize even complex patterns in a data set without *a priori* understating of the underlying mechanism. Hence, we have used ANNs to develop a model linking CH₄ emissions to five of the reservoir parameters about which data is most commonly available in the prior art. Using a compendium of all available data on these parameters, of which a small part was kept aside for use in model validation, it has been possible to develop a model which is able to forecast CH₄ emissions with a root mean square error of 37. It indicates a precision significantly better than the ones achieved in previous reports. The model provides a means to estimate CH₄ emissions from reservoirs of which age, mean depth, surface area, latitude and longitude are known.

Keywords: reservoirs; global warming; methane; modeling; artificial neural networks (ANNs)

1. Introduction

During the last three decades, human-made reservoirs have been increasingly implicated in contributing global warming gases (GHGs), carbon dioxide (CO₂), methane (CH₄), and nitrous oxide (N₂O) into the atmosphere [1,2]. More recently information is also emerging that shallow reservoirs created for small hydropower generation units and/or irrigation, may be contributing several times more GHGs, per unit surface area of the water-spread, than deeper reservoirs created for larger-scale use [3]. However, despite a very large number of artificial reservoirs functioning across the world, as also the spurt in the creation of small hydropower generation facilities on the presumption that they are cleaner than larger hydro [4,5], very few measurements have actually been done on CO₂ emissions from human-made reservoirs.

In recent reports, Deemer et al., [6]; the World Bank [1], and Chen et al. [7] have emphasized the paucity of information that exists on the emission of carbon dioxide from hydroelectric and other surface water reservoirs. This has been so despite the fact that the first report on this subject had come over a quarter-century ago—in 1993 [8]—and the issue has been hotly debated ever since, remaining alive in scientific circles, as well as layperson's concerns.

Chen et al., [7] have also brought out that attempts have been made from time to time to develop models with which measured carbon dioxide emissions can be correlated with factors like reservoir dissolved oxygen, temperature, pH, organic matter content, etc.—when known. The aim has been

to use such models to predict CO₂ emissions from other reservoirs where CO₂ emissions have not been measured. However, as detailed by Chen et al. [7], the models developed hitherto have been exceedingly short on precision. They are also constrained by regional limitations.

If the situation vis-à-vis CO₂ emissions from reservoirs is so restrictive, it is even more unfavorable when it comes to the other major greenhouse gas associated with the reservoirs—methane (CH₄). The hypolimnions of the reservoirs, in which dissolved oxygen levels can fall to zero near the water-sediment interface [9–11], can be major producers of methane. At times the methane production can be so high that it can become a source of energy [12,13] on being collected and purified [14]. Moreover, CH₄ has 34 and 86 times higher global warming potential (GWP) than CO₂, on 100-year and 20-year time horizons, respectively [14]. This can make its contribution to the overall greenhouse gas (GHG) emissions from reservoirs highly significant.

As reviewed by the World Bank [1], several authors have made estimates of CO₂ and CH₄ emissions from the earth's surface-water reservoirs. These estimates vary widely, by orders of magnitude—ranging from 3380 Tg CO₂-eq/year [15] to 209 Tg CO₂-eq/year [16]. The reasons behind a disparity of such a magnitude are a large number of factors that influence the GHG emissions of the reservoirs, as well as the great shortage and patchiness of the information available to gauge their relative contributions. As noted by the World Bank [1], if the earliest and the most liberal estimate of St Luis et al. [15] was plagued by uncertainties, so is the latest and the most conservative estimate of Prairie et al., [16].

The present work is a first-ever attempt to use the bioinspired artificial intelligence technique of artificial neural networks (ANN) in linking CH₄ emissions to some of the more commonly studied aspects of the world's reservoirs. ANNs have gained increasing popularity, due to their ability to model several types of complex phenomena which defy handling by analytical methods. This is made possible by the ability of the ANNs to sense patterns in data sets even under conditions of high imprecision and noise [17]. In situations where the cause-effect relationships may not be overt or their mechanism not established, ANNs can still model a phenomenon. However, despite the fact that the use of ANNs in several aspects of climate change studies has been extensive, only one past attempt at using ANNs for GHG emissions from reservoirs has been made so far by Chen et al., [7], and the study was limited to CO₂.

This paper demonstrates the applicability of multi-layer-perceptron artificial neural networks (MLP-ANN) for the prediction of CH₄ emission fluxes from reservoirs using latitude, longitude, age, mean depth and surface area as predictor input variables.

2. Methodology

2.1. Data Acquisition

The latest compendium which has included all the data generated prior to its publication, is by Deemer et al. [6]. We have added to it the information that has been reported since then up to the present. The resulting database covers 276 reservoirs from tropical, subtropical, temperate, Mediterranean and boreal regions of the world. The data encompasses morphometric, geographic, historical, physico-chemical and biological aspects along with information on emission of CO₂ and CH₄. However, these aspects have not been covered to equal extents for all the reservoirs. Indeed, only in very few cases have several of the above-mentioned aspects been covered. The CH₄ emissions have been given for only 167 of the reservoirs. We found that the aspects covered for the greatest number of reservoirs—92, for which CH₄ fluxes are also available, are latitude, longitude, age, mean depth and surface area. Hence, this data set was used for the ANN modelling.

2.2. Data Pre-Processing

To avoid overfitting, assignment of weighing factors that may be far off the mark, and the bias and convergence problems, it is necessary that the data being fed to the ANN is scaled to a uniform

range or normalized to have a mean value of zero and standard deviation of unity. Accordingly, in the present study, a typical ANN processing function provided by the neural network toolbox of MATLAB R 2017b, called *mapminmax*, was used to scale the data to the range -1 to 1 .

2.3. Network Architecture and Training

Multilayer networks are known to perform both linear and non-linear computations and are capable of producing good approximations of functions [18]. In view of this, the Levenberg-Marquardt (L-M) back-propagation algorithm has been used for ‘training’ the network. The training comprised of tuning the values of weights and biases of the layers by successive iterations. Before the training started, the network was fed with two sets of data. One, designated ‘input data’, consisted of input variables (reservoir age, mean depth, surface area, latitude and longitude) which we had acquired from the prior art as explained in Section 2.1. The other, termed ‘target data’ comprised of the output variable (methane flux) which had also been acquired from the prior art and which corresponded with the input data. The weights and biases of the ANN were altered iteratively until the network was able to forecast the output data corresponding to its input dataset achieving as close a match as possible between the output data and the target data. In other words, the training led to an ANN model which, upon feeding the input data, was able to generate output which matched most closely with the corresponding target data. This manner of training was done multiple times until the best fit was obtained. It must be noted that every time the network was put through training, its weights and biases were re-initialized.

We have chosen the L-M training algorithm also because several authors in the past have found it ideal [7,19], especially for small and medium-sized networks because system memory is not a constraint in dealing with such networks [18]. A three-layered feed-forward network was constructed which had a hyperbolic tangent-sigmoid transfer function (*tansig*) in the hidden layers and a linear transfer function (*purelin*) in the output layer. These functions were used to calculate a layer’s output from its net input:

$$tansig(A_j) = \frac{2}{(1 + e^{-2N_j}) - 1} \quad (1)$$

$$purelin(A_j) = N_j,$$

where $N_j = \sum_{i=1}^n W_{ij}$. $A_i + b_j$ is the sum of weighted inputs and biases for each neuron in the j th layer; n is the number of output layer neurons, W_{ij} is the weight of the neurons between i th and j th layers; A_i the i th neuron output, and b_j is the bias of the j th neuron.

The network architecture is shown in Figure 1. Its input layer had five neurons, the first hidden layer constituted 40 neurons, the second hidden layer had 30 neurons, and the output had one neuron. The number of layers and the neurons were selected on the basis of the configuration, which gave the best network performance. Four inputs were fitted to a single output.

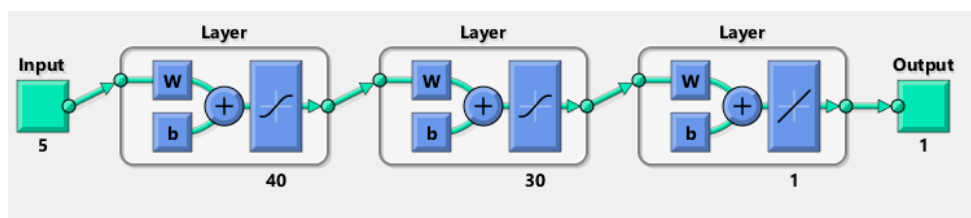


Figure 1. The architecture of the network employed by us.

The performance of the network was evaluated using mean square error (MSE) and regression analysis as indicators of the closeness of fit. The performance goal was set at 0.001. Usually, network training ceases when either the performance goal or the maximum epoch conditions are met; an epoch being the number of complete iterations consisting of output generation and error backpropagation.

The maximum number of iterations for this network (epochs) to converge the computation was set at 2000. The learning rate, which controls the degree/size of weight and bias changes during the network learning phase of the training algorithm, was set at 0.01.

3. Results and Discussion

3.1. Developing the ANN Model through Training

Out of the 92 data sets, acquired as detailed in Section 2.1, five data sets were randomly pulled out and were kept aside to test the network performance in making predictions in relation to the data not used in model development. The remaining 87 data sets were used in model development. The range and standard deviations of all the input and output parameters are given in Table 1.

Table 1. Range of variables in the data sets chosen for the study.

Aspect	Latitude	Longitude	Age (years)	Mean Depth (m)	Surface Area (km ²)	Methane Emissions (mg·CH ₄ ·m ⁻² ·day ⁻¹)
Number of samples	92	92	92	92	92	92
Minimum	−42.73	−126.00	0.80	0.50	0.005	0.0360
Maximum	68.00	146.17	104.00	158.10	4318.00	1719.54
Average	–	–	36.07 ± 25.22	29.26 ± 31.5	464.77 ± 767.86	81.49 ± 255.35

3.2. Model Performance

The optimal network configuration which gave the best performance was selected. The performance was analyzed by plotting a regression line between the network output and the observed methane fluxes in the training set (targets). The network layer weights and biases were re-initialized by training the network again and again, until the best fit was obtained. This training ceased when the MSE reached 1325.54 at epoch 1623, after which there was no further decrease in the MSE.

The RMSE and the mean absolute error (MAE) were seen to be 36.4 and 12.1, respectively. The best fit gave an R^2 value of 0.98 (correlation coefficient of 0.99) as indicated in Figure 2. These compare very favorably with the performance of the Levenberg-Marquardt backpropagation, neural network (LM-BPNN) model on CO₂ emissions of Chen et al., [7], which is the only model other than these authors', that has been reported so far on GHG emissions from surface water reservoirs. The RMSE (396.6) of the model of Chen et al., [7] was over 10 times greater, while the MAE (268.5) was over 20 times greater. On the other hand, the coefficient of determination of 0.52 between the observed and the predicted values of the model of Chen et al. [7] was about half. The closeness of the predicted values with that of the actual methane emissions can be seen in Figure 3.

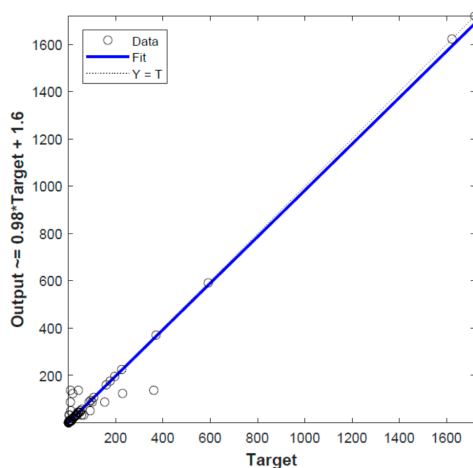


Figure 2. Regression plot of the modelled network output versus target data.

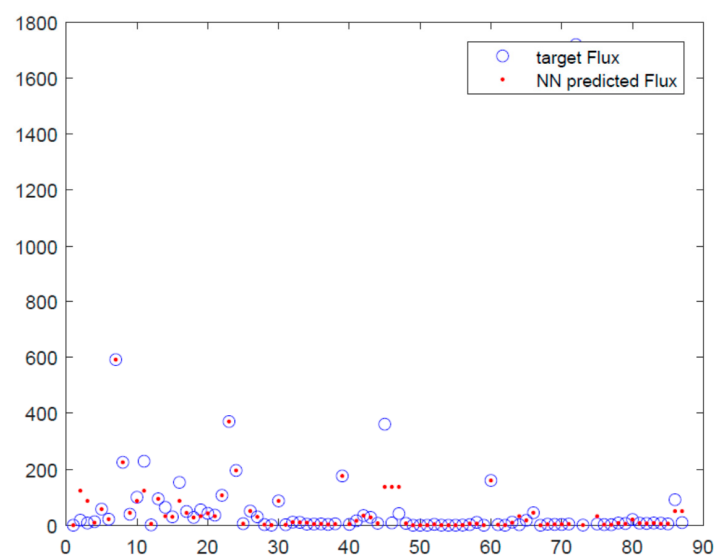


Figure 3. Comparison of observed methane flux (target) values with model-predicted values for the training data set.

The absolute error between the model output and the actual flux for each reservoir lied between -215 to 119.8 , as indicated by the error histogram in Figure 4. However, most of the errors fell between -39 to 49 , with 71% of the instances lying in the close range -3.5 to $+0.5$.

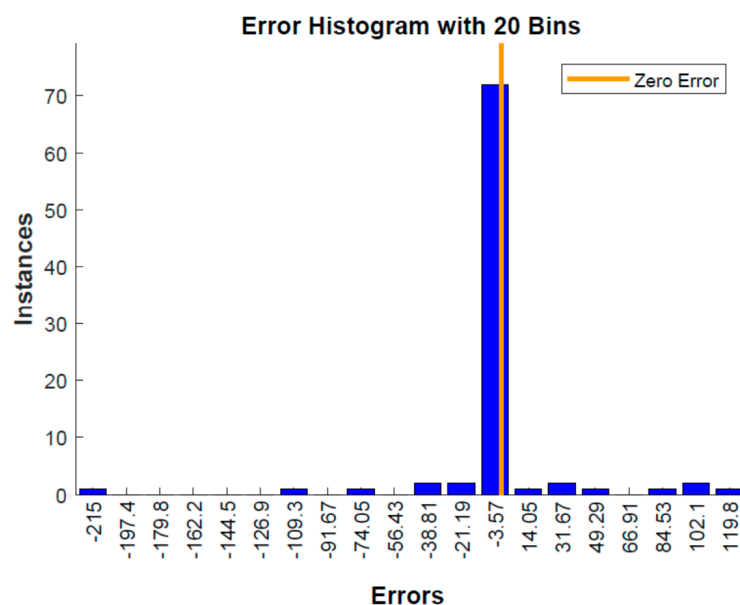


Figure 4. Error histogram of the network.

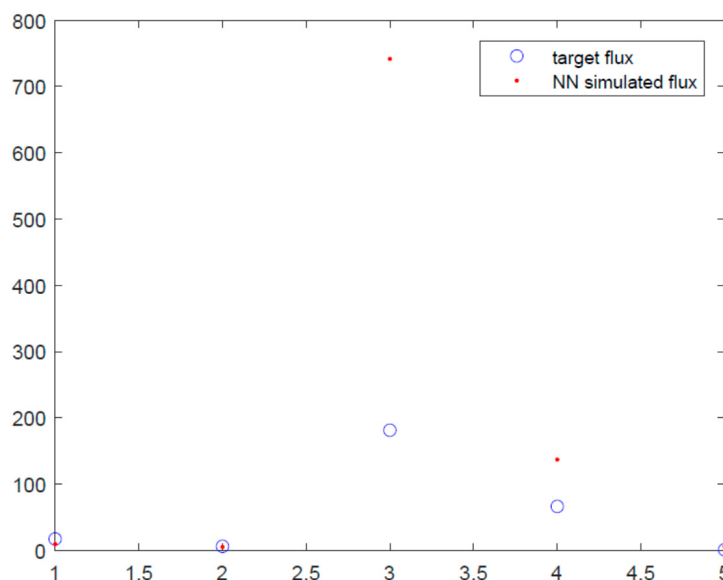
3.3. Simulation

The ANN model was tested to determine its ability to predict methane flux values from experimentally generated data which had not already been used in the training process. In other words, the ANN model was made to forecast methane flux values from a set of inputs that had not been used in model development. The input data used for the simulation, the actual flux and the flux values simulated by the model are given in Table 2.

Table 2. A comparison between the actual and simulated methane emissions.

Reservoir Identifier	Climate	Latitude	Longitude	Age (years)	Mean Depth (m)	Surface Area (km ²)	Measured Methane Flux (mg·CH ₄ ·C·m ⁻² ·day ⁻¹)	ANN Simulated Methane Flux (mg·CH ₄ ·C·m ⁻² ·day ⁻¹)
Hartwell	Temperate	34.47	−82.85	50	13.9	226	17.09	9.27
La Grande 3	Boreal	54.2	−76.13	19	24.8	2420	6.075	5.03
Wilcza Wola	Subtropical	21.9	50.35	21	2.6	1.6	181.26	741.82
Lacamas Lake	Temperate	45.37	−122.25	73	7.8	1.3	66.27	136.94
Manic 3	Boreal	50.23	−69.03	28	44.1	236	0.825	5.93

A correlation of 0.98 (R^2 of 0.968) was obtained between the simulated flux versus the actual flux. Figure 5 gives the comparison of the actual methane flux value with the model simulated values. The RMSE is 252.7, and the MAE is 129.04. Here, again, the present model for CH₄ has performed well in comparison to the model of Chen et al., [7] for CO₂. Its RMSE, MAE and R^2 values are half, one-third and half, respectively, of the corresponding statistic of RMSE of 505.43, MAE of 395.33 and R^2 of 0.47 reported by Chen et al. [7] during the testing phase.

**Figure 5.** Comparison of the actual methane flux with its model simulated values.

3.4. Applications

Latitude and longitude are indirect indicators of the climate under which different surface water reservoirs exist. Along with the reservoir age, mean depth and surface area, the climate (especially ambient temperature) is the main factor that influences CH₄ emissions, because temperatures of 35 ± 2 °C foster methanogenesis, while lower temperatures slow it down [20,21]. Younger reservoirs tend to emit more CH₄ than older reservoirs because the former have larger stocks of organic carbon acquired from the vegetation that was submerged in the course of the filling of the reservoir [22,23]. As these stocks dwindle over the years, methane generation from this source decreases even if not in proportion because organic carbon keeps coming into the reservoir, due to the gradual shift of the reservoir from oligotrophy to mesotrophy and then eutrophy [8]. It is not uncommon to find shallower areas of the reservoirs infested with aquatic weeds and/or algae. Upon senescence, their biomass sinks to the bottom and gets fermented in the anoxic zone normally existing in the hypolimnion in water layers close to the reservoir bottom [24].

Reservoir depths also influence CH₄ emissions by the role they play in influencing the depths of the anoxic zones and in the manner and place of the release of CH₄. In deep reservoirs, the hypolimnion is significantly colder than the epilimnion and dissolves much larger quantities of CH₄ than shallower

reservoirs can do. When the hypolimnic water is released, and its temperature gradually increases downstream, the dissolved CH_4 is let off [13].

The water spread area influences the organic carbon production, which occurs in the photic zone, mainly in the epilimnion, and larger the reservoir area more epilimnion space it makes available for primary production. Shallow reservoirs, which have large area-volume ratios, can contribute several times more GHGs than deeper reservoirs of lower area-volume ratios [3].

The model developed in this work makes use of these five parameters. They are all easy to measure. Even though primary literature carries only 96 datasets which cover these five parameters (along with CH_4 flux), information on these five parameters is likely to be available with the agencies managing the reservoirs in different parts of the world. Once that data is acquired, it can enable the prediction of CH_4 emissions from all those reservoirs. The model reported here can, thus, be used to get reasonable estimates of global, regional and local contributions of the reservoirs to CH_4 emissions without the need for expensive experimentation.

4. Conclusions

The paper has successfully demonstrated the use of artificial neural networks (ANN)—based on multi-layer perceptron, and trained using the Levenberg-Marquardt (L-M) back-propagation algorithm—to model methane emission fluxes from human-made water reservoirs. A three-layered network architecture with five input neurons, 40 and 30 neurons in the two hidden layers, and one output neuron, gave the best results. The model was trained using data from 87 reservoirs distributed in tropical, subtropical, temperate, Mediterranean and boreal regions of the world. The model's RMSE came to 36.4, and the regression coefficient R^2 between the model-predicted and actual flux was 0.98. The model was then used to simulate methane fluxes from five independent reservoirs not used in the training phase. It led to an R^2 of 0.968 (correlation of 0.98) between the simulated and the actual fluxes, an RMSE of 252.7 and a mean absolute error of 129. These findings indicate that the bioinspired artificial intelligence technique of ANN can be successfully employed in predicting methane fluxes from reservoirs using input indicators such as reservoir age, mean depth, surface area and latitude. These input indicators are likely to be available (even though they may not yet be in the public domain) for most reservoirs of the world. The model, thus, makes it possible to estimate CH_4 fluxes of all those reservoirs. In situations where data on reservoir surface area and depth is available at corresponding reservoir age, the model can forecast the changes in the CH_4 fluxes over time. Moreover, if a systematic study is planned in future, in which approximately 100 reservoirs are studied in terms of a larger number of parameters that influence CH_4 emission, it should be possible to use that data in developing more powerful models based on ANNs, with greater predictive ability.

Author Contributions: Conceptualization, T.-A., S.A.A. and T.A.; methodology, T.-A., S.A.A., T.A.; software, T.-A., T.A.; validation, T.-A., T.A.; formal analysis, T.-A.; investigation, C.L.; resources, S.A.A.; data curation, C.L.; writing—original draft preparation, S.A.A.; writing—review and editing, S.A.A.; visualization, T.-A.; supervision, S.A.A.; project administration, S.A.A.; funding acquisition, T.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: S. A. Abbasi thanks the Council of Scientific and Industrial Research (CSIR), New Delhi, for the Emeritus Scientist grant (21(1034)/16/EMR-II). Tabassum-Abbasi thanks the authorities of the University of Petroleum and Energy Studies for encouragement.

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

References

1. World Bank. *Greenhouse Gases from Reservoirs Caused by Biogeochemical Processes*; World Bank: Washington, DC, USA, 2017.
2. Abbasi, S.A.; Abbasi, T. *Current Concerns in Environmental Engineering*; Nova Science Publishers: New York, NY, USA, 2019; p. 309. ISBN 978-1-53613-920-4.

3. Ollivier, Q.R.; Maher, D.T.; Pitfield, C.; Macreadie, P.I. Punching above their weight: Large release of greenhouse gases from small agricultural dams. *Global Chang. Biol.* **2018**, *25*, 721–732. [[CrossRef](#)] [[PubMed](#)]
4. Abbasi, T.; Abbasi, S.A. *Remote Sensing, GIS and Wetland Management*; Discovery Publishing House: New Delhi, India, 2010; 411p.
5. Premalatha, M.; Tabassum-Abbasi; Abbasi, S.A. A critical view on the eco-friendliness of small and large hydroelectric installations. *Sci. Total Environ.* **2014**, *481*, 638–643. [[CrossRef](#)] [[PubMed](#)]
6. Deemer, B.R.; Harrison, J.A.; Li, S.; Beaulieu, J.J.; DelSontro, T.; Barros, N.; Bezerra-Neto, J.F.; Powers, S.M.; Dos Santos, M.A.; Vonk, J.A. Greenhouse gas emissions from reservoir water surfaces: A new global synthesis. *BioScience* **2016**, *66*, 949–964. [[CrossRef](#)]
7. Chen, Z.; Ye, X.; Huang, P. Estimating carbon dioxide (CO₂) emissions from reservoirs using artificial neural networks. *Water* **2018**, *10*, 26. [[CrossRef](#)]
8. Rudd, J.W.M.; Hecky, R.E. Are hydroelectric reservoirs significant sources of greenhouse gases? *Ambio* **1993**, *22*, 246–248.
9. Abbasi, S.A. *Wetlands of India, Lakes of Peninsular India*; Discovery Publishing House: New Delhi, India, 1997; Volume 1, 151p.
10. Chari, K.B.; Richa Sharma.; Abbasi, S.A. *Comprehensive Environmental Impact Assessment of Water Resources Projects*; Discovery Publishing House: New Delhi, India, 2005; Volume 1, 580p.
11. Ganesh, P.S.; Ramasamy, E.V.; Gajalakshmi, S.; Abbasi, S.A. Extraction of volatile fatty acids (VFAs) from water hyacinth using inexpensive contraptions, and the use of the VFAs as feed supplement in conventional digesters. *Biochem. Eng. J.* **2005**, *27*, 17–23. [[CrossRef](#)]
12. Bambace, L.A.W.; Ramos, F.M.; Lima, I.B.T.; Rosa, R. Mitigation and recovery of methane emission from tropical hydroelectric dams. *Energy* **2007**, *32*, 1038–1046. [[CrossRef](#)]
13. Ramos, F.M.; Bambace, L.A.W.; Lima, I.B.T.; Rosa, R.R.; Mazzi, E.A.; Fearnside, P.M. Methane stocks in tropical hydropower reservoirs as a potential energy source. *Clim. Chang.* **2009**, *93*, 1–13. [[CrossRef](#)]
14. Baena-Moreno, F.M.; Rodríguez-Galán, M.; Vega, F.; Vilches, L.F.; Navarrete, B. Recent advances in biogas purifying technologies. *Int. J. Green Energy* **2019**, *16*, 401–412. [[CrossRef](#)]
15. St. Louis, V.L.; Kelly, C.A.; Duchemin, É.; Rudd, J.W.; Rosenberg, D.M. Reservoir surfaces as sources of greenhouse gases to the atmosphere: A global estimate. *BioScience* **2000**, *50*, 766–775. [[CrossRef](#)]
16. Prairie, Y.T.; Alm, J.; Beaulieu, J.; Barros, N.; Battin, T.; Cole, J.; del Giorgio, P.; DelSontro, T.; Guérin, F.; Harby, A.; et al. GHG emissions from freshwater reservoirs: What does the atmosphere see? *Ecosystems* **2017**, *21*, 1058–1071. [[CrossRef](#)] [[PubMed](#)]
17. Mosavi, A.; Salimi, M.; Faizollahzadeh, Ardabili, S.; Rabczuk, T.; Shamshirband, S.; Varkonyi-Koczy, A.R. State of the art of machine learning models in energy systems, a systematic review. *Energies* **2019**, *12*, 1301. [[CrossRef](#)]
18. Mathworks. *Global Optimization Toolbox: User's Guide (R2017b)*; The MathWorks, Inc.: Natick, MA, USA, 2017.
19. He, B.; Oki, T.; Sun, F.B.; Komori, D.; Kanae, S.; Wang, Y.; Kim, H.; Yamazaki, D. Estimating monthly total nitrogen concentration in streams by using artificial neural network. *J. Environ. Manag.* **2011**, *92*, 172–177. [[CrossRef](#)] [[PubMed](#)]
20. Abbasi, T.; Tauseef, S.M.; Abbasi, S.A. Anaerobic digestion for global warming control and energy generation: An overview. *Renew. Sustain. Energy Rev.* **2012**, *16*, 3228–3242. [[CrossRef](#)]
21. Abbasi, T.; Tauseef, S.M.; Abbasi, S.A. *Biogas Energy*; Springer: New York, NY, USA; London, UK, 2012; 169p, ISBN 978-1-4614-1039-3.
22. Ramos, F.M.; Lima, I.B.; Rosa, R.R.; Mazzi, E.A.; Carvalho, J.C.; Rasera, M.F.; Ometto, J.P.; Assireu, A.T.; Stech, J.L. Extreme event dynamics in methane ebullition fluxes from tropical reservoirs. *Geophys. Res. Lett.* **2006**, *33*, 1–4. [[CrossRef](#)]
23. Lima, I.B.; Ramos, F.M.; Bambace, L.A.; Rosa, R.R. Methane emissions from large dams as renewable energy resources: A developing nation perspective. *Mitig. Adapt. Strateg. Glob. Chang.* **2008**, *13*, 193–206. [[CrossRef](#)]
24. Chari, K.B.; Abbasi, S.A. Wetlands—An overview. *Hydrol. J.* **2003**, *26*, 41–58.

