Optimization of the selection process of the co-substrates for chicken manure fermentation using neural modeling

Andrzej Lewicki[1], Artur Bugała[2], Hanna Piekarska-Boniecka[3],Aleksandra Lewicka[4], Piotr Boniecki[1], Marta Cieślik[1]

[1] *Poznan University of Life Sciences, Institute of Biosystem Engineering, ul. WojskaPolskiego 28, 60-637 Poznan, Poland*
[2]*Poznan University of Technology, Institute of Electrical Engineering and Electronics, ul. Piotrowo 3a, 60-965Poznan, Poland*
[3]*Poznan University of Life Sciences, Department of Entomology and Environmental Protection, ul. WojskaPolskiego 28, 60-637 Poznan, Poland*
[4]*Poznan University of Technology,Institute of Environmental Engineering,ul. Piotrowo 3a, 60-965 Poznan, Poland*

Abstract. Intense development of research equipment leads directly to increasing cognitive abilities. However, along with the raising amount of data generated, the development of the techniques allowing the analysis is also essential. Currently, one of the most dynamically developing branch of computer science and mathematics are the Artificial Neural Networks (ANN). Their main advantage is very high ability to solve the regression and approximation issues. This paper presents the possibility of application of artificial intelligence methods to optimize the selection of co-substrates intended for methane fermentation of chicken manure. 4-layer MLP network has proven to be the optimal structure modeling the obtained empirical data.

1.Introduction

ArtificialNeural Networks(ANN) are an importantapplicationof cognitivemethodsusedin the areaof empirical researchcarried out inthe area of widelyunderstoodartificial intelligence[1-4].In particular,promising results are related to an application one of the importantcharacteristics ofANN, i.e. the ability to solve the problems concerning regressionandapproximation issues. It givesthe possibility of using of efficientANNsimulators,among others, in order toshape up the predictiveissues. ANNcanbesuccessfully used practically in anysituation, where themain objective is toestimatethe value ofa variablebased onacquired(e.g. experimentally) characteristics(i.e. descriptors) [5-7].

However, inavailable literature related to the discussed topic there is little informationconcerning the use ofANNtomodelthe process of biomethane emission[8, 9].It seems highly appropriateto tryto buildthe regressionneural model, generatedbased onempirical datacollectedduringresearch conductedunder laboratory conditions.Developed andtestedneural modelcanserve as a toolsupporting thedecision-making processesthat occur during operationof biogas plants[10].Itscorrect applicationimproves and rationalizes thesystem forselecting theproper mixture ofinput substratesso that themethane fermentation processtakes placeunderpossibly optimal conditions.

The objective of the paper was todevelopthe neuralestimatorintendedto predictthe amount ofemittedbiomethanefrom the fermentation process ofchicken manurewithadditions of othersubstrates.As atraining setwas usedthe databaseof the biogasefficiency ofthe substrates andtheirmixtures obtainedin the Laboratory ofEcotechnologyat the PoznanUniversity of Life Sciencesunder the scientific supervisionof Prof.JacekDach.

The Laboratory ofEcotechnologyis the largestPolishbiogas laboratory, withmore than250fermentersoperating accordingto DIN38414/S8**[**11].

2.Methodology

In order to accomplishthe aforementioned task, the followingworkflow has beenproposedandthen implemented(Fig. 1).

Figure 1. Action scheme

The most important step in ANN generating is to produce an adequate training set containing in its structure encoded empirical data. For this purpose, the Authors have defined the numeric input variables and predicted output variable resulted from the structure of the formulated scientific problem. Five input variables have been adopted as a share of respondents substrates:

- GRASS - grass [g Fresh Mass (FM)/reactor],
- STRAW - straw [g FM/reactor],
- CHICKEN - chicken manure [g FM/reactor],
- MAIZE - maize silage [g FM/reactor],
- INOCULU - inoculum [g FM/reactor].
  Jako 1 zmienną wyjściową przyjęto:
- BIOMETH - biomethaneproduction efficiency [m3/Mg FM].

Using theobtainedexperimental results,set of empirical dataconsistingof 156measurement cases has been generated. The above mentioned set has beenused tocreateANNand, therefore, has been dividedinto 3subsets:

- training fileincluding 78 cases,
- validationfileincluding39 cases,
- test fileincluding39 cases.

Some part oftraining file(cases:33to 39) intended for ANN generatoris showninTable 1.

Table 1. Fragment of the training file for the ANN simulator (BIOMETH – output variable)

No.	GRASS	STRAW	CHICKEN	MAIZE	INOCULU	**BIOMETH**
33	0.00	13.55	0.00	0.00	1192	**180.89**
34	0.00	13.55	0.00	0.00	1210	**175.45**
35	000	0.00	38.85	0.00	1164	**70.51**
36	0.00	0.00	38.95	0.00	1162	**68.19**
37	0.00	0.00	38.80	0.00	1161	**69.35**
38	0.00	0.00	0.00	43.	1156	**109.19**
39	0.00	0.00	0.00	43.	1156	**113.14**

The simulator of artificial neural networksimplementedin a commercialStatisticapackagehas been usedfor the design ofneural models. The following types ofneural networkshave been subjected totesting:

- linear networks.
- MLP networks (MultiLayer Perceptron).
- RBF networks (Radial Basic Function).
- GRNN networks (Generalized Regression Neural Network – regressive networks).

The development of neural modelswas a 2-stage procedure.At the beginning,an effectiveoption supporting the process of ANN designinghas been used, i.e. *AutomatedWebDesigner*implementedin Statistica computer system. This toolallowed toautomateandsimplify theprocedures of preliminaryseeking of a set with predictiveneural networksmodeling the investigated process.
In the second stage of neural models development another useful device has been used i.e. *User Networks Designer*. It offers the possibility of advanced interference in the parameters and training methods of generated neural networks. This tool has been activated frequently in order to modify both the initial settings of parameters, learning algorithms and the ANN structure itself.

3.Results

3.1 Artificial neural networks

From 100of generatedneural models,a fileof10isolatedneuraltopologies has been extracted as it showsTable 2 (ANN no. 10is the bestnetwork).

Table 2. File of 10 generated ANN

	Type	RMS error	Imputs	Layer1	Layer2	Quality
1	RBF	11.953	5	9	-	0.29790
2	MLP	11.406	5	1	-	0.26561
3	RBF	9.8235	5	10	-	0.24572
4	RBF	9.1699	5	11	-	0.22898
5	MLP	8.5263	5	2	-	0.21456
6	RBF	5.3826	5	13	-	0.13413
7	MLP	4.4734	5	3	-	0.11134
8	MLP	3.7940	5	6	-	0.09538
9	RBF	3.5926	5	14	-	0.08896
10	**MLP**	**2.655757**	**5**	**8**	**4**	**0.0633**

where:
- RMS (Root Mean Square) error - it is a total error made by the network on a data set (learning, test or validation). It is determined throughout summing the squared individual errors, then dividing the obtained sum by the number of included values and determining the square root of the quotient obtained. RMS error is the most convenient single value to be interpreted that describes the total ANN error.
- Qality – the value of root mean square error regression.

4-layer MLP networkwith structureshown in Figure 1turned out to be the optimalstructure modelingobtainedempirical data. The input layeris composed offiveneuronswitha linearPSP(Postsynaptic Function)andactivation function. The firsthidden layeris composed oftwosigmoidalneurons, i.e. with linearPSPfunction andlogisticactivation function. The secondhidden layeris built fromthreeneuronswith structureidentical to theneurons from thefirst layer. The networkoutputwasonesigmoidal neuron. The developedneural modelwastrainedusingBP(BackPropagation) method in 3 cyclesfor 50epochsandoptimized with algorithmCG(Conjugate Gradients) for 196epochs. The following parameters have been adopted during the training process with the algorithm of BP error:

- decreasing training coefficient: $\eta = 0.1$ do $\eta = 0.01$,
- momentumfactor: $\alpha = 0.22$.

The structure ofgeneratedANN, typeMLP: 5-8-4-1,is shown in Fig.2.

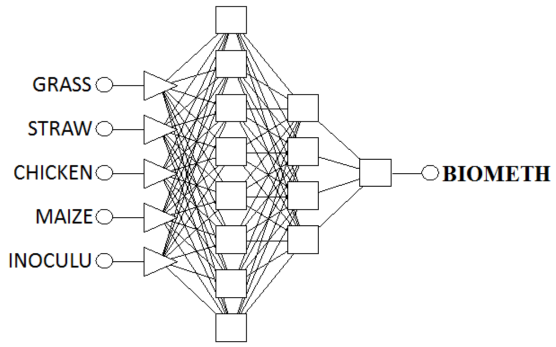

Figure 2.Structure ofthe optimalANN, typeMLP: 5-8-4-1, with2hidden layers.

Unidirectional MLP neural networks are commonly used in ANN topology practice. Multilayer Perceptron represents the so-called category of parametric neural models. Where characteristic is that the number of neurons constituting its structure is considerably less than the number of cases of the training set. The basic characteristics of MLP network include the following features:

- MLP is a unidirectional network,
- MLP is trained by „with-a-teacher" method,
- has a multi-layer structure, with the following layers: input, hidden, output
- architecture of the connections within the network allows communication only between the neurons located in contiguous layers,
- neurons being a part of ANN, MLP type,aggregate the inputdataby defining theinputs weightedsums(using the linearformula ofaggregation),
- activation functionof the inputneuronsis linear, hidden neurons-non-linear, whilethe nature ofoutput neuronsis generallynonlinear,
- due to thesaturation level present(in the sigmoidactivationfunctions), all the data processed by the networkrequire an appropriaterescaling(preprocessing andpost processing).

The quality of the generated MLP network, as a predictive tool, is identified by the so-called statistics of regression issues, which are shown in Table 3.

Table 3. Regression statistics of generated model for the files: training. validation and test

	biomethane for training file	**biomethane** for validation file	**biomethane** for test file
Data mean	70,81500	79,05513	89,66154
Data S.D.	45,39954	52,06174	40,76000
Error mean	0,442888	-0,07916	0,304094
Error S.D.	3,655942	4,18808	3,908609
Abs E. Mean	2,672447	2,971817	2,857903
S.D. Ratio	**0,080530**	**0,080440**	**0,095890**
Correlation	**0,996799**	**0,996759**	**0,995508**

where:

- Data mean - the mean value of the output variable, calculated on the basis of the preset values of this variable, collected (respectively) in the training, validation or testing set. The regression statistics are calculated independently for the training, validation and test set.
- Data S.D. - the standard deviation calculated for the specified (as above) values of the output variable.
- Error mean - mean error (module of the difference between the reference value and the value obtained at the output) for the output variable.
- Error S.D. - the standard deviation of the errors for the output variable.
- Abs E. Mean - mean absolute error (the difference between the reference value and the value obtained at the output) for the output variable.
- S.D. Ratio - quotient of the standard deviations both for errors and data. This is the main indicator of the quality of the regression model developed by the network.
- Correlation - standard R.Pearson correlation coefficient for the setpoint value and the value obtained at the output.

Table 3shows thatthe correlation isat the level of0.99for the following files:training, validationandtest one, while the quotient of standard deviations for errorsand thedatarangesfrom 0.08in case ofvalidation file up to0.09forthe test file.

The assessment of sensitivity of developed MLP on individual input variables has been performed in order to determine the level of significance of representative parameters used to build the neural model. The procedure of sensitivity analysis is implemented in the Statistica package as a tool for assessing the impact of the various input variables on the quality performance of generated neural model. The sensitivity analysis provides an insight into the usefulness of particular input variables. Moreover, it indicates high-ranking variables that without any loss of quality of the network can be omitted. Furthermore it also points the key variables (low rank value), which must not be ignored.

Table 4. The values of quotients of errors and rank for the five input variables

	grass	**straw**	chicken	**maize**
Rank	**3**	**1**	4	**2**
Error	17.69924	67.48457	17.64613	32.38631
Ratio	4.614127	17.59298	4.600281	8.442992

gdzie:

- Rank - significance level of input variable, organizes the variables by importance: 1 is the dominant variable.
- Error - the network quality in the absence of a variable: the lower the rank number of the ANN input variable, the higher error of reduced network (without the input variable).
- Ratio - quotient of the error of the reduced network by the ANN error obtained using all variables. If the Ratio is less than one, than removal of the variable improves the quality of the network.

The sensitivity analysis of the MLPneural model5-8-4-1onthe input variables of analyzed processshowed thatthe most importantparametersin the process of neural estimation of the amount ofgeneratedbiomethaneare(in order):

- Rank 1: STRAW
- Rank 2: MAIZE
- Rank 3: GRASS

In order to visualize the behavior of the generated neural model, depending on the values of the main descriptors (STRAW, MAIZE, GRASS) Figure 3 shows the three surfaces of ANN responds illustrating the biomethane efficiency in a function of key input variables of developed neural model.

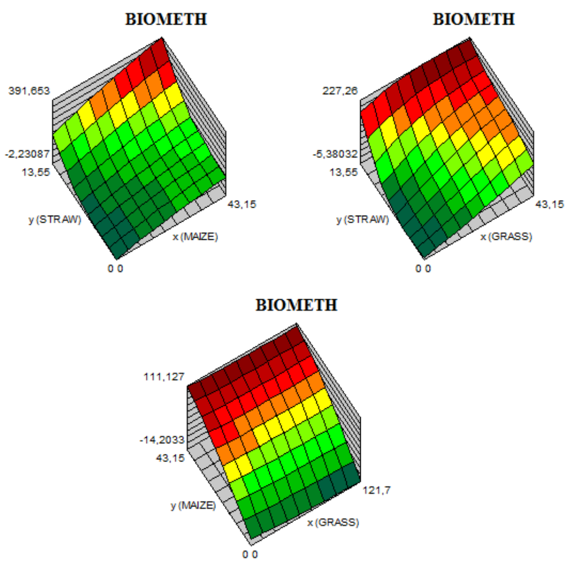

Figure 3. Answer surfacesfor network MLP:5-8-4-1

4.Discussion

The analysis of the empirical resultsobtainedduring the testsor receivedin the course of industrial processes controlsometimesis very complex.Usuallyit is related to a largeamount ofdata obtained.
Duringcontinuous processessuch asindustrial production ofbiogasthroughoutmethane fermentationwe have to dealwith a specific case.In the aforementioned situationit is possible toobtain a veryextensivetraining setbased ona number of variablesanalyzedover a long periodof time[12,13].In order to properlyinterpret the multifactorialresultsit is necessary torefer toadvancedstatistical methodsbased onArtificial Intelligence.According to[14; 15**]** application of thistype of analysis enables defining the trends, dependencies and propercontrol of the processes,not only in thelaboratory scale butmainlyin the industrial one.

5. Conclusions

The resultsanalyzedusingANNhave shownthat the optimumstructuremodelingobtainedempirical dataisa 4-layer MLP network. The receivedvisualizationsand statementsare consistentwith the experience ofa biogas plantstaffandscientists studyingthe efficiencyof biogassubstrates.It proves that chosen method can besuccessfully implemented into the plannedapplication-optimizingthe selection of the co-substratesfor fermentation of the chicken manurein order to obtainthe highest possibleof biomethane production.

Acknowledgement

This paper was created as a result of the realization of the project entitled “The innovative energy production in biogas plant through utilization of poultry manure with the alternation of a plant substrate for algae”financed by The National Centre for Research and Development (contract no.: GEKON1/O5/214528/34/2015).

References

1. Czekała W., Dach J., Ludwiczak A., Przybylak A., Boniecki P., Koszela K., Zaborowicz M., Przybył K., Wojcieszak D., Witaszek K.. *Proc. SPIE. 9631, Seventh International Conference on Digital Image Processing (ICDIP 2015),* **963117**. (doi: 10.1117/12.2197041 (2015)
2. Przybył K., Koszela K., Boniecki P., Mueller W., Raba B., Lewicki A. Proc. SPIE **9159**, *Sixth International Conference on Digital Image Processing (ICDIP 2014),* **91590W**; doi:10.1117/12.2064243 (2014)
3. Raba B., Nowakowski K., Lewicki A., Przybył K., Zaborowicz M., Koszela K., Boniecki P. SPIE 9159,

Sixth International Conference on Digital Image Processing (ICDIP 2014), **91590H** (April 16, 2014); doi:10.1117/12.2064091 (2014)

4. Dach J., Koszela K., Boniecki P., Zaborowicz M., Lewicki A,., Czekała W., Skwarcz J., Qiao W., Piekarska-Boniecka H., Białobrzewski I. Renewable andSustainable Energy Reviews**56**, 603–610 (2016)
5. Boniecki P., Piekarska-Boniecka H., Koszela K., Zaborowicz M., Przybył K., Wojcieszak D., Zbytek Z., Ludwiczak. A., Przybylak A. , Lewicki A. *Proc. SPIE. 9631, Seventh International Conference on Digital Image Processing* (ICDIP 2015), **963115**. (July 06, 2015) doi: 10.1117/12.2197029 (2015)
6. Koszela K; J. Otrząsek; M. Zaborowicz; P. Boniecki; W. Mueller; B. Raba; A. Lewicki; K. Przybył. Proc. SPIE 9159, *Sixth International Conference on Digital Image Processing (ICDIP 2014),* **915913**; doi:10.1117/12.2064274 (2014)
7. Kujawa S., Tomczak R.J., Kluza T., Weres J., Boniecki P.,. *4th International Conference on Digital Image Processing. Kuala Lumpur (Malaysia), proceedings of SPIE,* Vol. **8334**, Article Number: 83341R, DOI: 10.1117/12.949899 (2012)
8. GueguimKanaa E.B., Olokeb J.K., Lateefb A., Adesiyanb M.O. Renewable Energy, Volume **46**, Pages 276–281
9. Lewicki A., Boniecki P., Dach J., Janczak D., Czekała W., Zbytek Z., Cieślik M.. *4TH International Conference on Materials Engineering for Advanced Technologies (ICMEAT 2015)* Pages: 649-653 (2015)
10. Cerbin S., Nowakowski K., Dach J., Pilarski K., Boniecki P., Przybyl J., Lewicki A., *4th International Conference On Digital Image Processing (ICDIP 2012). Proceedings of SPIE.* Vol.: **8334** Article Number: 83342A Doi: 10.1117/12.954164 (2012)
11. Janczak D., Marciniak M., Lewicki A., Czekała W., Witaszek K., Rodríguez Carmona P. C., Cieślik M., Dach J., *Procedia Technology,* Vol. **8**, 209–214. DOI: 10.1016/j.protcy.2013.11.029 (2013)
12. Li J., Wei L., Duan Q., Hu G., Zhang G. Bioresource Technology, Volume **156**, Pages 307–313 (2014)
13. Li J., Kong C., Duan Q., Luo T., Mei Z., Lei Y. Bioresource Technology, Volume **193**, Pages 62–67 (2015)
14. Ozkaya B., Demir A., Sinan Bilgili M. Environmental Modelling & Software, Volume 22, Issue **6**, Pages 815–822 (2007)
15. Abu Qdaisa H., BaniHanib K., Shatnawia N. Resources, Conservation and Recycling, Volume **54**, Issue 6, Pages 359–363 (2010)