Alicja KOLASA-WIĘCEK
Opole University of Technology, Department of Economics and Regional Research, Faculty of Economy and Management
Waryńskiego 4, 45-047 Opole
e-mail: a.kolasa-wiecek@po.opole.pl

RECURSIVE PARTITIONING APPLICATION IN THE ASSESSMENT OF THE CLIMATIC CONDITIONS IMPACT OF NON- CO_2 GHGS ON AGRICULTURAL EMISSIONS

Summary

Agricultural practices contribute to emissions of the greenhouse gases (GHGs) such a methan (CH_4) and nitrous oxide (N_2O). Their estimated share from agricultural sources is assessed at around 50% of CH_4 and 60% of N_2O emissions. The efforts made by the agricultural sector aim to limit and reduce emissions. Due to their significant share, all the complementary knowledge information concerning their reduction are highly precious. The paper proposes the use of neural modeling techniques and the summary of results by modeling based on a decision tree (Recursive Partitioning) to estimate the levels of methane and nitrous oxide emissions from agriculture under varying weather conditions in Poland. The obtained results support the hypothesis that neural model describing the effect of meteorological conditions on the $CH_4$and N_2O emissions is an appropriate tool for the assessment of the projected emission level.
Key words: *greenhouse gases, CH_4, N_2O, modeling, neural networks, decision trees, meteorological conditions*

ZASTOSOWANIE METODY RECURSIVE PARTITIONING W OCENIE WPŁYWU WARUNKÓW KLIMATYCZNYCH NA ROLNICZE EMISJE GAZÓW CIEPLARNIANYCH INNYCH NIŻ CO_2

Streszczenie

Praktyki rolnicze przyczyniają się do emisji gazów cieplarnianych (GGC), takich jak metan (CH_4) i podtlenku azotu (N_2O). Ich szacunkowy udział ze źródeł rolniczych oceniany jest na około 50% emisji CH_4 i 60% emisji N_2O. Wysiłki podejmowane przez sektor rolny mają na celu ograniczenie i redukcję ich emisji. Ze względu na ich znaczący udział, wszelkie informacje dopełniające wiedzę na temat możliwości ich redukcji są niezwykle cenne. W pracy zaproponowano wykorzystanie technik neuronowego modelowania oraz posumowania wyników z wykorzystaniem modelowania w oparciu o drzewo decyzyjne (Recursive Partitioning) do estymacji poziomu metanu i podtlenku azotu emitowanych z rolnictwa przy zmiennych warunkach meteorologicznych w Polsce. Uzyskane wyniki badań potwierdzają hipotezę, że model neuronowy, opisujący wpływ warunków meteorologicznych na emisję CH_4 i N_2O, jest właściwym instrumentem dla dokonania oceny prognozowania poziomu tej emisji.
Słowa kluczowe: *gazy cieplarniane, CH_4, N_2O, modelowanie, sieci neuronowe, drzewa decyzyjne, warunki meteorologiczne*

1. Introduction

Agricultural practices contribute to emissions of the greenhouse gases methane (CH_4) and nitrous oxide (N_2O). According to the United Nations Framework Convention on Climate Change (UNFCCC) and Kyoto Protocol under it, industrial countries have to estimate their greenhouse gas emissions annually, and assess the uncertainties in these estimates. N_2O and CH_4 are potent greenhouse gases with relative global warming potentials of 310 and 21 times those of carbon dioxide (CO_2) on a mole per mole basis [10, 26].

The calculations prove that the IPCC, CH_4 and N_2O from agricultural sources make up about 50% and near 60% of total anthropogenic emissions, appropriately [11] with a broad variety of uncertainty in the estimates of both the agricultural contribution and the anthropogenic total.

Annual non-CO_2 emissions from agriculture are expected to increase in the coming decades because of food demands and shift in diet [20].

It is expected that agricultural N_2O emissions will increase by 50% by 2020 due to increased use of nitrogen fertilizers and enhanced animal manure production (relative to 1990) [6].

If the increase of the number of livestock will continue, the CH_4 emissions will increase proportionally as projected at around 60% by 2030 [13, 28].

Therefore, it is so important to reduce emissions of greenhouse gases from agricultural sources. The global technical mitigation potential from agriculture by 2030, considering all gases was estimated to approximately 4,500 [3] to 5,500-6,000 Mt CO_2 –eq/yr [20].

The mitigation of GHG emissions in agriculture is characterized by a range of options. Improved crop and grazing land management are among the most prominent choices (such as improved agronomic practices, nutrient use, tillage, and residue management), as well as restoration of organic soils that are drained for crop production and restoration of degraded lands. A significant, but a bit lower mitigation is also possible with improved water and rice management, set-asides, land use change (e.g., transformation of cropland to grassland) and agro-forestry; as well as improved livestock and manure management.

Among the factors which have an impact on the amount of greenhouse gas emissions generated by animals, among others are: the species, the type of fodder, the kind of breeding and the environmental conditions [18].

Some studies, carried out in a narrow area of selected aspects suggest that the weather conditions, especially temperature and precipitation, may affect the size of the released CH_4 and N_2O emissions [21].

In other studies a noticeable effect of both tillage practices and weather on the amount of released greenhouse gases was seen [1].

Obligations that Poland has taken by joining the ranks of the European Union, prompt to the necessity to seek new complementary knowledge information in the reduction of GHGs emissions, in various sectors of the economy. Food safety is a significant issue, but its activities have important effects on the environment quality.

Efforts undertaken by the agricultural sector aim at the limit and reduction of non-CO_2 greenhouse gas emissions. In this context, taking new exertions to obtain complementary knowledge information about the mechanisms/ conditions affecting the quantity of emissions released, seem to be very reasonable.

2. The aim of the study

Aim of the study is to assess the impact of meteorological conditions on the rate of the non-CO_2 greenhouse gases released from agriculture. The paper assumed that the use of neural network modelling techniques to estimate emissions and modelling using Recursive Partitioning classification tree are appropriate instruments.

3. Methodology research

The input material for agricultural emissions were data coming from international statistics database of the UNFCCC [31] and the meteorological conditions derived from the IMGW (Institute of Meteorology and Water Management) in Warsaw. Meteorological data came from monitoring stations deployed throughout the country (Gdynia, Legionowo, Warszawa-Bielany, Radzyń, Jarczew, Puławy, Kołobrzeg, Łeba, Piła, Toruń, Mikołajki, Gorzów Wielkopolski, Koło, Legnica, Wieluń, Łódź - Lublinek, Sulejów, Włodawa, Jelenia Góra, Śnieżka, Kłodzko, Bielsko-Biała - Aleksandrowice, Zakopane, Kasprowy Wierch, Olesko).

The analysis includes the average annual of the following parameters: temperature, precipitation, insolation, humidity and wind velocity. The study period accepted for the analysis is dictated by number of available data. UNFCCC database records cover a period of 20 years of greenhouse gas emissions, therefore, such a period was adopted for the study. Due to the relatively small amount of data was decided to choose Bayesian networks, which can be safely used for a limited range of training data [16]. The study was conducted using a Flexible Bayesian Models on Neural Networks, Gaussian Processes, and Mixtures.

These networks are increasingly and successfully used in a wide spectrum of sciences being currently of great interest to the world of science and indicating credibility and trust that are put in them [4, 5, 7, 14, 19, 24, 29].

Flexible Bayesian regression models are the base of Flexible Network Bayesian Neural Networks and multilayer perceptron neural networks or Gaussian processes are the basis for classification applications. The Markov chain Monte Carlo method is used in the network (MCMC). Markov chain sampling is supported by software modules which are included in the distribution, and can be useful in other applications. Bayesian networks make use of complex neural network models without concern for "overfitting" that can occur with traditional methods of learning neural networks.

Bayesian network learning methods provoked controversy for decades [15]. In recent years they are in a position of great trust [25].

The basic Bayesian model has two levels of specification reliability function

$$O \mid \theta \sim f(O\ \theta) \tag{1}$$

and *a priori* specification

$$\theta \sim \pi(\theta) \tag{2}$$

where O i θ can be vectors.

In the simplest Bayesian analysis it is assumed that π is known and after a transformation with the function of the credibility, the distribution *a posterior* θ can be presented in a form:

$$p(\theta|O) = \frac{f(O\ |\theta)\pi(\theta)}{\int f(O\ |\theta)\pi(\theta)d\theta} = \frac{f(O\ |\theta)\pi(\theta)}{m(O)} \tag{3}$$

where the integration results in the denominator m (O) is the boundary density of the data set O.

This equation is a general form of Bayes' theorem, where the knowledge about the experimental data (represented by the credibility function f) with *a priori* opinion (expressed by π distribution) are linked, to form a probability distribution *a posteriori* (determined by p). Taking into account the boundary distribution pattern assumes an abbreviated formula:

$$p\ (\theta\ |O) = f(O\ |\theta)\ \pi(\theta) \tag{4}$$

specifying that the *a posteriori* distribution is proportional to the product of the credibility function and the *a priori* distribution [24, 29].

All the variables are subjected to modelling random variables, size, parameters and probability distributions, and the inference is carried out by constructing *a posteriori* probability distribution for observed and unobserved volume of interest on the basis of the observed sample given in the information as well as *a priori* assumptions.

In standard neural network techniques, the average predictions for specific and desirable complexity of the model are strict and often in terms of computational time-consuming and labor intensive. [2]. In the case of the Bayesian approach the solution of this issue is more natural and logical. The unknown level of complexity is determined by defining the so-called uninformed a priori assumptions for the hyperparameters which define the model complexity. The obtained model predictions are averaged and weighted against all elements of its complexity through probabilities distribution *a posteriori* assigned with a sample of taken information. Such a model also allows for existing a various complexity of its different parts by grouping network parameters which are interchangeable and have the same role in the model and common hyperparameters [12].

Summaries of prediction results were performed on the basis of a Recursive partitioning decision tree . The algorithm was created in an software environment R v. 2.10.0, implemented in *rpart* package (Fig. 1) and the result is a graph in the form of a classification tree.

```
> read.table("C:/Documents and Settings/ala/Moje dokumenty/publikacje/art. 2013/dane.klimat/RPART/klimat.txt",header=T)->klimat
> attach(klimat)
> require(rpart)
Loading required package: rpart
> fit=rpart(V6~V1+V2+V3+V4+V5)
> plot(fit);text(fit)
```

Fig. 1. Loading a data set with a function call *rpart* in an R programme
Rys. 1. Wczytanie zbioru danych wraz z wywołaniem funkcji rpart w środowisku R

Recursive partitioning methods are popular and widely used tools for nonparametric regression and classification in many scientific fields [8, 22, 23]. The applications of these methods are far reaching. [9]. It produces a classification tree in which subjects are assigned to mutually exclusive subsets according to a set of predictor variables [17]. The basic steps of a proposed algorithm for model-based recursive partitioning are: (1) adapt a parametric model to a dataset; (2) check for parameter instability over a set of partitioning variables; (3) if there is some overall parameter instability, divide the model with respect to the variable associated with the highest instability; (4) recur the procedure in each of the daughter nodes.

The statistics of the analyzed parameters are presented in Table 1.

4. Results and discussion

Neural analysis was carried out using a Flexible Bayesian Models on Neural Networks, Gaussian Processes, and Mixtures [Neal 2000], working on Unix/Linux. In a carried prediction, for parameters with varying meteorological conditions, expected CH_4 and N_2O emissions were proposed. Two separate series of tests for CH_4 and N_2O emissions were performed.

Selected basic numerical and graphical quality parameters of learning the network, among others the so called recoil index and the trajectory graph of the control values, called hyperparameters weight, provide the correct and optimal process or learning network. The results of learning neural network, for example one of the cases, shown in Figure 2, coefficient obtained at the level 0.412, ie within the limits of variation from 0.2 to 0.8, and relatively parallel trajectories hyperparameters weight, can attest to achieve a balance in the impulses flow through the network, thereby the reliability of prediction (the details of neural analysis – see: [16].

In neural prediction a number of combinations were assumed. Due to the fact of the scale of the obtained results, a classification tree was used to make it possible to summarize them. The most important results of the analysis of Recursive Partitioning are illustrated in Figures 3 and 4.

Table 1. Statistical description of the variables; own calculations
Tabela 2. Statystyczny opis zmiennych; obliczenia własne

Variable	Min	Mean	Median	Standard deviation	Max
Insolation	4.208997	4.642357	4.62649	0.247395	5.186883
Humidity	72.50437	78.59785	78.9366	2.086107	81.08531
Temperature	6.367188	8.217917	8.154119	0.702875	9.277188
Precipitation	460.3013	563.8837	560.5312	63.67116	730.324
Velocity of wind	2.794965	3.032104	3.016181	0.167204	3.397708

The analyzed variables are listed in a form of percentiles (Table 2).

Table 2. Percentiles values for attributes; own calculations
Tabela 2. Wartości percentyli atrybutów; obliczenia własne

Percentiles	Insolation, h	Humidity, %	Temperature, °C	Precipitation, mm	Velocity of wind, m/s
10%	4.3499	77.45447	7.515278	500.3414	2.808767
20%	4.404556	78.01057	7.671667	510.0114	2.887617
30%	4.496111	78.08465	7.934219	520.3348	2.916615
40%	4.624233	78.27067	8.071406	550.3805	2.991406
50%	4.62649	78.9366	8.154119	560.5312	3.016181
60%	4.719645	79.42688	8.38413	560.9768	3.058021
70%	4.752207	79.70776	8.575521	600.3471	3.132118
80%	4.785569	80.0773	8.902865	610.7172	3.178958
90%	4.963221	80.73118	9.087847	620.8201	3.234722

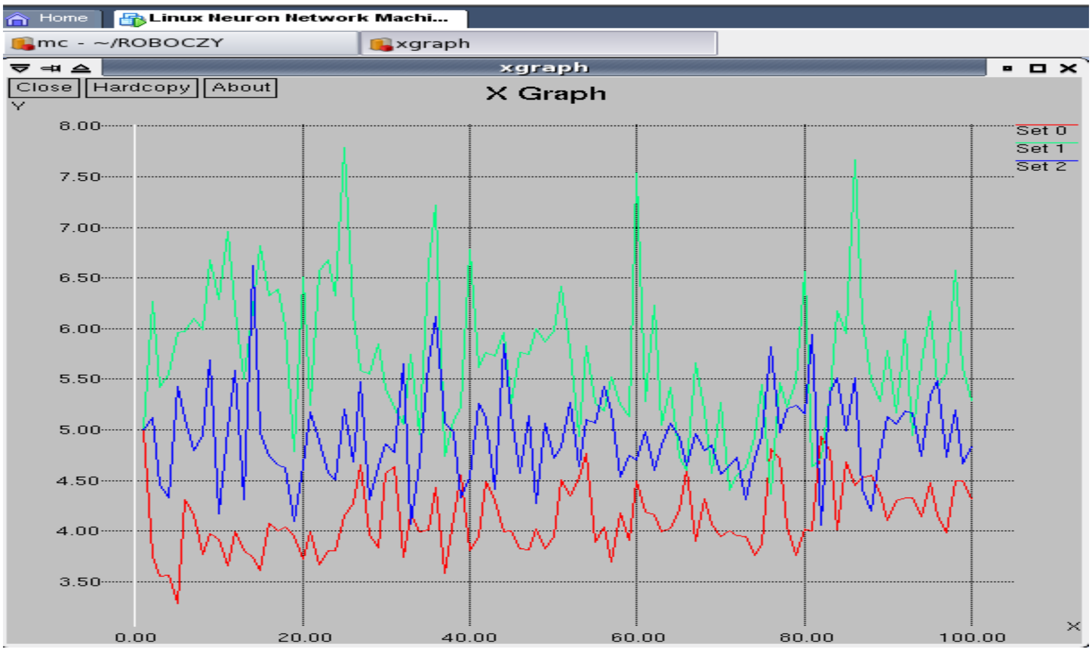

Fig. 2. Selected graphic parameters for neural network learning in the FBM application
Rys. 2. Wybrane graficzne parametry uczenia się sieci neuronowej w programie FBM

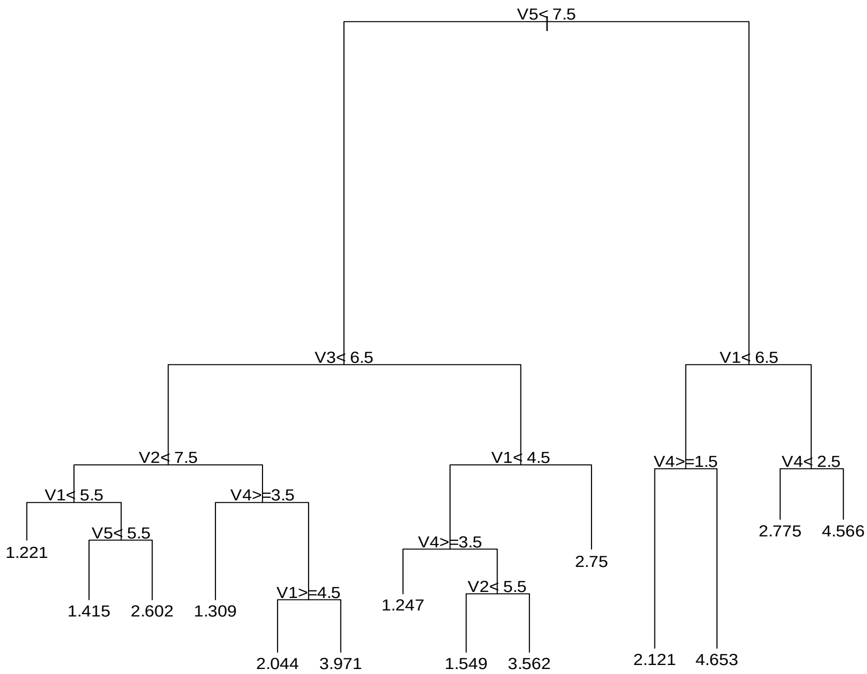

Fig. 3. Recursive Partitioning graph for CH_4 emissions depending on meteorological conditions, where: V1 – insolation, V2- humidity, V3 – temperature, V4 – precipitation, V5 – velocity of wind; own calculations
Rys. 3. Graf drzewa Recursive Partitioning dla emisji CH_4 w zależności od warunków meteorologicznych, gdzie: V1 – nasłonecznienie, V2- wilgotność, V3 – temperatura, V4 – opady, V5 – prędkość wiatru; obliczenia własne

Analysis of the classification tree for CH_4 emissions leads to the conclusion that the most important variable is the wind velocity (Fig. 3). On the graph presenting a wind velocity above 3.18 m/s, the highest CH_4 emissions are noted. Furthermore, in the case of the highest emission a significant role was observed when the insolation reaches a value below 4.72 h, simultaneously with a very low precipitation – beginning from 505.2 mm and below.

On the side of the graph which corresponds to the wind velocity less than 3.13 m / s, the lowest CH_4emissions value are found. For achieving this scale, meteorological conditions are as follows: the average annual temperature is lower than the (V3 <6.5) 8.38 ° C, humidity level (V2 <7.5) less than 79.70%, while the insolation (V1 <5.5) less than 4.63 h per day. A very similar minimum value of the of CH_4 emissions were also noticed with the increase of temperature - from 8.57°C, with insolation below 4.62 h/day, and the precipitation beginning from 535.4 mm and above.

Figure 4 presents the results of analysis of the Recursive Partitioning for N_2O emissions.

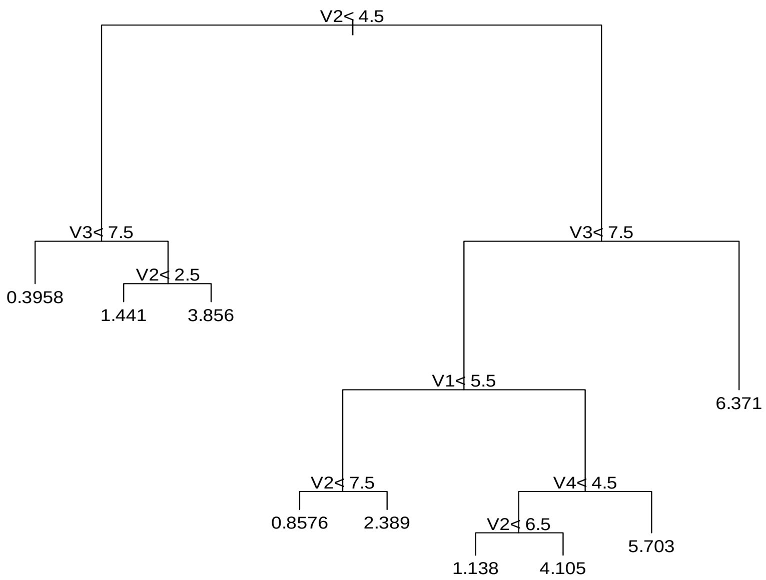

Fig. 4. Recursive Partitioning graph for N_2O emissions depending on meteorological conditions, where: V1 – insolation, V2- humidity, V3 – temperature, V4 – precipitation, V5 – velocity of wind; own calculations

Rys. 4. Graf drzewa Recursive Partitioning dla emisji N_2O w zależności od warunków meteorologicznych, gdzie: V1 – nasłonecznienie, V2- wilgotność, V3 – temperatura, V4 – opady, V5 – prędkość wiatru; obliczenia własne

With regard to N_2O emission the analysis by a rpart method gives a tree structure with a less extensive structure in comparison to the example considered above. Generally it can be concluded that the two factors, ie humidity and temperature will play a key role in N_2O emissions.

The analysis leads to the following conclusions: variable V2- humidity separates the set of records by the value 4.5. For the humidity level below 78.27% and temperatures below 8.57°C, the lowest N_2O emissions were noted. The right side of the graph, with the humidity value from 78.94% and a simultaneous increase in temperatures - beginning from 8.9°C illustrates conditions for which the highest N_2O emissions were achieved.

5. Conclusions

The results achieved through the classification tree Recursive Partitioning illustrate conditions for which pollutants CH_4 and N_2O may reach the lowest and the highest emissions.

The study results confirm the hypothesis that modelling with the use of the Recursive Partitioning method, describing the CH_4 and N_2O emission with variable meteorological conditions, it is an appropriate instrument for the assessment of the level of prediction emissions.

The analysis prompts to the conclusion that the consideration of meteorological conditions on the basis of national data would allow for increasing the accuracy of estimates of greenhouse gases emissions from agriculture.

6. References

[1] Ball B.C., Scott A., Parker J.P.: Field N_2O, CO_2 and CH_4 fluxes in relation to tillage. Compaction and soil quality in Scotland. Soil & Tillage Research, 53: 29-39, 1999.

[2] Buntine W.L., Weingend A.S.: Bayesian back-propogation. Complex system, 5, 6: 603-643, 1991.

[3] Caldeira K., Morgan M.G., Baldocci D., Brewer P.G., Chen C.T.A., Nabuurs G.J., Nakicenovic N., Robertson G.P.: A portfolio of carbon management options. In: The Global Carbon Cycle: Integrating Humans, Climate, and the Natural World, Eds. Field, C.B., Raupach M.R. Island Press, Washington, DC, 103-129, 2004.

[4] Dose V., Menzel A.: Bayesian analysis of climate change impacts in phenology. Global Change Biology, 10, 2: 259-272, 2004.

[5] Ellison A.M.. Bayesian inference in ecology. Ecology Latters, 7, 6: 509-2520, 2004.

[6] Food and Agricultural Organization, 2003: World Agriculture: Towards 2015/2030. An FAO Perspective, FAO, Rome 2003.

[7] Furrer R., Sain S.R., Nychka D., Meehl G.A:. Multivariate Bayesian Analysis in atmosphere – ocean general circulation models. Environmental and Ecological Statistics, 14, 3: 249-266, 2007.

[8] Heping Z., Burton H.S.: Recursive Partitioning and Applications, Springer-Verlag Gmbh. 2010.

[9] Heping Z.: Recursive Partitioning and Tree-based Methods. Springer Handbooks of Computational Statistics, 853-882, 2012.

[10] IPCC. Radiative Forcing of Climate Change. The 1996 Report of the Scientific Assessment Working Group of IPCC Summary for Policy Makers. World Meteorology Organization. UN Environment Program, Geneva, Switzerland, 1996.

[11] IPCC. Climate change 2001. The scientific basis–contribution of work group I to the third assessment report of IPCC. Cambridge University Press, Cambridge, 2001.

[12] Lampinen J., Vehtari A.: Bayesian approach for neural networks-review and case studies, Neural Networks, 14: 257-274, 2001.

[13] Mosier A., Kroeze C.: Potential impact on the global atmospheric N_2O budget of the increased nitrogen input required to meet future global food demands. Chemosphere-Global Change Science, 2: 465-473, 2000.

[14] Nazir N., Mir A.H., A.A. Khan A.A: Bayesian analysis of mixed effect models and its applications in agriculture, Electronic Journal of Applied Statistical Analysis, 4, 2: 164 – 171, 2011.

[15] Neal R.: Bayesian Learning for Neural Networks, Springer – Verlag GnbH, 1996.

[16] Neal R.: Flexible Bayesian Models on Neural Networks, Gaussian Processes and Mixtures v. 2004-11-10, University of Toronto, 2004.
[17] Nelson L.M., Bloch D.A., Longstreth W.T. Jr., Shi H.: Recursive partitioning for the identification of disease risk subgroups: a case-control study of subarachnoid hemorrhage, Journal of Clinical Epidemiology, 51, 3: 199-209, 1998.
[18] Pasyniuk P.: Reducing greenhouse gas emissions by replacing diesel by vegetable oils. Problems of Agricultural Engineering, 4: 79-89, 2010.
[19] Sadoddin A., Shahabi M.: Agricultural drought management using a Bayesian decision model for rainfed wheat farmlands in east of Golestan Province, Iran, Workshop on metrics and methodologies of estimation of extreme climate event, UNESCO, Paris, 2010.
[20] Smith P., Martino D., Cai Z., Gwary D., Janzen H., Kumar P., McCarl B., Ogle S., O'Mara F., Rice C., Scholes B., Sirotenko O.: Agriculture. In Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Eds. B. Metz, O.R. Davidson, P.R. Bosch, R. Dave, L.A. Meyer, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.
[21] Song C.C., Zhang L.H., Wang Y.Y., Zhao Z.C.: Annual dynamics of CO_2, CH_4, N_2O emissions from freshwater marshes and affected by nitrogen fertilization, Huan Jing Ke Xue, 27, 12: 2369-2375, 2006.
[22] Strobl C., Malley J., Tutz G.: An introduction to recursive partitioning: rationale. Application. and characteristics of classification and regression trees. Bagging. and random forests. Psychological Methods, 14, 4: 323-48. 2009.
[23] Su X., Tsai Ch., Wang H., Nickerson D.M., Li B.: Subgroup Analysis via Recursive Partitioning. Journal of Machine Learning Research, 10: 141-158, 2009.
[24] Tukiendorf A.: Selected aspects of contemporary spatial epidemiology of cancer example in Opole province, Institute of Occupational Medicine named prof named J. Nofera, 2004, [in polish].
[25] Tukiendorf M.: Neural Modeling of mixing of heterogeneous granular systems. Dissertations of Agricultural University in Lublin, 272, 2003, [in polish].
[26] Ullah S., Frasier R., King L., Picotte-Anderson N., Moore T.R.: Potential fluxes of N_2O and CH_4 from soils of three forest types in Eastern Canada. Soil Biology and Biochemistry, 40, 986–994, 2008.
[27] Unsitalo L.: Advantages and challenges of Bayesian networks in environmental modelling. Ecological Modelling, 203, 3-4: 2312-318, 2007.
[28] United States Environmental Protection Agency (USEPA): Global Anthropogenic Non-CO2 Greenhouse Gas Emissions: 1990- 2020. EPA, 2006. Washington, http://www.epa.gov/non-co2/econinv/downloads/GlobalAnthroEmissionsReport.pdf [Accessed: 3.02.2013].
[29] Whithers S.D.: Quantitative methods: Bayesian inference, Bayesian thinking, Progress in Human Geography, 26, 4: 553-566, 2002.
[30] Zeileis A., Hothorn T., Hornik K.: Model-Based Recursive Partitioning. Journal of Computational and Graphical Statistics, 17, 2: 492-514. 2008.
[31] United Nations Framework Convention on Climate Change: http://unfccc.int/di/DetailedByGas/Event.do?event=go [Accessed: 10.12.2012]