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

USE OF THE CLASSIFICATION TREE MODELING TO INVESTIGATE THE INFLUENCE OF CROPS ON N_2O AND CH_4 EMISSIONS RELEASED FROM THE AGRICULTURAL SECTOR

Summary

Methane and nitrous oxide are key pollutants emitted from agriculture. Primarily the livestock production has a significant share in CH_4 emissions. The N_2O emissions largely correspond to direct emissions associated with the cultivation of soils. The priority task of agriculture is to develop adaptive solutions enabling the reduction of pollutions in the next years. These capabilities apply to both technological solutions on the farms, as well as improved methods of management and policy tools. Therefore complementary information to the knowledge in the field of the possibilities for reducing CH_4 and N_2O are extremely valuable. The study of predictions of N_2O and CH_4 emissions on the basis of different arable crops areas with the use of Flexible Bayesian Models of neural networks was carried out. The decision trees have been designed in order to provide the knowledge and methods that allow the rapid identification of the most important arable crops that affect the quantity of these emissions. On the basis of the conducted analysis, wheat, maize and potatoes in the case of N_2O emission and wheat and maize in the case of CH_4 emission are the most important differentiating variables.
Key words: *nitrous oxide, methane, crops, modeling, predicting, artificial neural network, classification tree*

1. Introduction

In the course of the broadly understood agricultural production a number of pollutants arises. The main sources of greenhouse gas emissions from the agricultural sector are usually considered to be direct emissions from crop and livestock production as well as changes of soil use and very rarely the use of means of transport [20, 24]. The key pollutants are methane and nitrous oxide which present a real threat to the environment. Efforts undertaken by the agricultural sector aim to limit and reduce these emissions, and represent one of the priorities of the modern food economy. One-third of N_2O emissions has anthropogenic nature [27]. Some of them are direct and other indirect emissions. Among direct sources it is possible to distinguish N_2O which is emitted directly to the atmosphere from agricultural land and fertilized and/or grazed pasture systems. Indirect emissions are caused by transfer of N from agricultural systems into soil and surface waters through drainage and surface runoff, or emission as ammonia or nitrogen oxides and deposition elsewhere, inducing N_2O production [22]. A by-product of microbial nitrification and denitrification processes utilizing manure nitrogen is also produced in manure as a result of N_2O emissions from Manure Management N_2O [14, 21]. Such factors as oxygen concentration, temperature, pH and microbial community affect the amount of N_2O produced which depends on the manure nitrification process. More than 71% of N_2O emissions associated with the direct cultivation of the soil, and nearly 26% of the emissions come from indirect processes - such as nitrogen dispersed into the environment at various stages of production in the field and on the farm [14]. Only about 2% of the emissions is attributed to animal faeces left on pastures [1].

The first complete inventory of N_2O emissions from agricultural sources in Poland with the application of a modified method of the IPCC and having taken into account the realities of Polish agriculture was performed at the Department of Chemistry of the Soil and Water in the Institute for Land Reclamation and Grassland Farming in 2002 [19]. There are large discrepancies in the literature in the assessment of N_2O emissions from agriculture areas in Poland [3]. NO_x emissions models from soils applied in the world are used for global inventory of N_2O emissions, but none of them is used throughout the world [21].

In Poland cereals are the leading group in arable crops. A large concentration of the crops is unfavorable for several reasons: it leads to yields reduction and increased use both of mineral fertilizers and plant protection products [11]. The basic kinds of plants cultivated in Poland include: wheat, rye, triticale, barley, oats, maize, oilseed rape, sugar beet and potatoes. The last two decades of arable crops are presented in Figure 1.

Agriculture, particularly livestock production, is a significant contributor to CH_4 emissions. The main sources include enteric fermentation and manure management. The digestive system of livestock produces CH_4 as a by-product of microbial fermentation of feed and is breathed out by the animals [7]. The main sources of CH_4 through enteric fermentation are both ruminants (cattle, sheep, goats) and non-ruminant livestock (horses and swine) which produce CH_4 in their digestion track. Different factors such as, ie. the type and amount of feed influences the amount of CH_4 produced [7, 15]. The decomposition of organic material in anaerobic conditions by facultative bacteria is the result of CH_4 production in livestock manure [9]. The amount of CH_4 produced is dependent on the amount of manure produced, manure characteristics (pH, nutrient content, volatile solids (VS) content), climatic conditions (temperature, rainfall) and the manure management system used [9, 15]. Handling manure in a solid form (e.g., in stacks) or depositing it on pasture, range, or paddock lands, it tends to decompose aerobically and produce little or no CH_4. The amount of CH_4 produced is affected by ambient temperature, moisture, and manure storage or residency time because they influence the growth of the bacteria responsible for CH_4 formation. Agricultural emissions of N_2O and CH_4 in last two decades in Poland presents Figure 2.

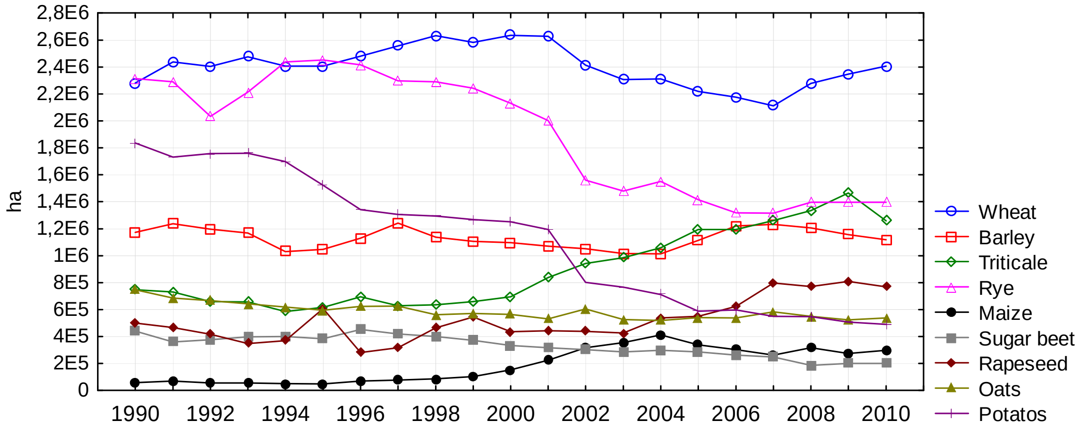

Fig. 1.The areas of the main crops in Poland [29]

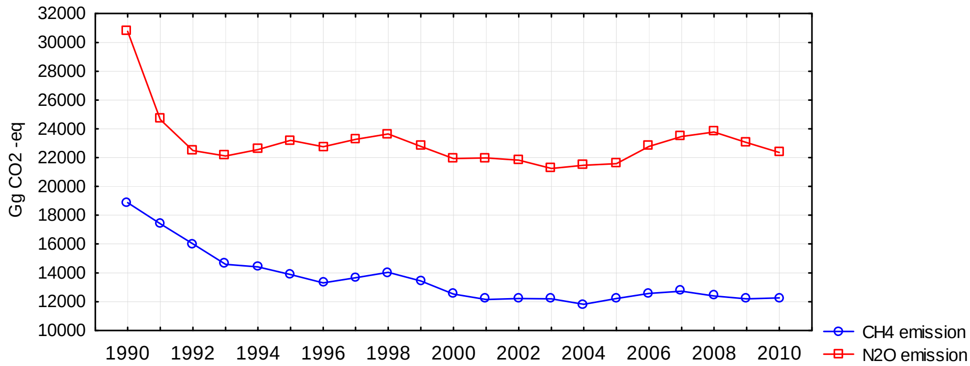

Fig. 2. Agricultural emissions of N_2O and CH_4 in Poland [30]

The priority agriculture mission in the coming years is to develop adaptive solutions which allow the pollutions reduction. These opportunities apply to both technological solutions on the farms, as well as improved methods of management and policy tools. Therefore complementary information to the knowledge of the possibility of reducing N_2O and CH_4 are extremely valuable.

2. The aim of the study

Each plant species has different nutritional requirements, as evidenced by the diverse potential yield and the percentage of individual components in the parts of the plants reaped from the field. So far proposed calculation methods of N_2O does not include the influence on emissions of the type of arable crops.

The effect of arable crops species on CH_4 emissions has not been previously recognized, although the type of feed consumed by the animals has an impact on the emissions of this pollutant [18]. It is estimated that the share of CH_4 emissions in burning of agricultural waste is different for each plant species [17].

Undertaken scientific problem was brought to:

1. The prediction of agricultural N_2O and CH_4 emissions on the basis of different area sizes of arable crops with the use of Flexible Bayesian Models neural networks,
2. Indicating variables values using the method of a decision tree (crop sizes) in which the emitted N_2O and CH_4 pollutants achieve the lowest and highest rates.

3. Methodology research

The study used data from international databases in the previous 20 years. Data concerning crops areas was derived from the FAO [29] and emissions data come from the UNFCCC database [30]. To start with, the importance of the attributes based on Spearman's test was verified. The selected variables were used for further analysis. They are summarized in Table 1.

Table 1. Statistical description of the variables; own calculations

Variable	Min	Mean	Median	Standard deviation	Max
Wheat	2,111,980	2,404,516	2,406,790	149,394.2	2,635,100
Rye	1,316,200	1,902,075	2,033,910	440,458.3	2,451,600
Maize	48,165	187,230	152,273	127,324.3	411,704
Sugar beet	187,484	329,775.1	333,131	80,311.28	452,638
Potatos	490,853	1,120,315	1,250,620	486,180.1	1,835,340

The variables are presented as percentiles (Table 2).

Table 2. Percentile values for analyzed attributes; own calculations

Percentiles	Wheat	Rye	Maize	Sugar beet	Potatos
	ha				
10%	2,218,090	1,395,600	54,521	199,936	548,884
20%	2,280,650	1,396,520	59,021	262,046	588,184
30%	2,310,740	1,479,300	70,123	286,300	713,250
40%	2,405,100	1,560,000	85,180	303,000	803,384
50%	2,406,790	2,033,910	152,273	333,131	1,250,620
60%	2,414,180	2,212,770	261,975	371,714	1,295,010
70%	2,476,870	2,289,500	298,700	384,477	1,341,890
80%	2,555,090	2,297,920	317,193	400,274	1,697,250
90%	2,627,050	2,414,980	339,342	419,364	1,757,340

The research was carried out with the use of Flexible Bayesian Models on Neural Networks [16] and R-Project package. The synthesis of the results was explored on the basis of the decision trees, which are one of the possibilities of constructing models. Classification tree modelling is a tool successfully used to solve a wide variety of problems [4, 5, 10, 13, 23, 25]. Trees are an effective method for scoop data and often used in the prediction. A decision tree is a directed acyclic graph. A rule for forecasting the class of an object from the values of its predictor variables is a classification tree. The tree is created by multiple dividing a learning sample of data in which the class label and the values of the predictor variables for each case are known. A node in the tree represents every partition [26]. It consists of a root and branches leading to the consecutive nodes. The root focuses the entire chosen sample. In this case, the classifier is represented by a binary tree, where nodes are questions about the values of the determined attribute and in the leaves are located the classes assessments [2, 12]. There are several reasons for the classification trees being especially attractive in a data mining environment. Not only, due to their intuitive representation, is the resulting classification model easy to assimilate by humans, but also classification trees are non-parametric. As no assumption about the underlying distribution is made by classification tree construction algorithms, they are especially suited for exploratory knowledge discovery. Moreover, classification trees can be constructed comparatively fast compared to other methods and their accuracy is comparable or superior to other classification models [6, 8].

The final result of analysis should be a tree with the least number of branches and nodes in order to determine the classification rules as simple as possible. In the R package the method of determining decision trees is available for instance as a function of *tree(tree)*, which was used in these analyses [2].

4. Results and discussion

Knowledge obtained from the selected (numerical and graphic) parameters of the Flexible Bayesian Models network learning, among others, so called recoil index and the trajectory graph of the control values, it is possible to draw conclusions the process of the network learning moved in optimal conditions. The balance in the impulses flow through the network that provides the resulting coefficient – 0.489 within the range of variability 0.2-0.8 proves that the network generalizes the acquired knowledge correctly.

Because of the numerous group of results (15 thousand. combinations), the synthesis of projected emissions are presented in the form of a classification tree. The results of a built tree for N_2O emissions are 10 divisions visible (Figure 3).

Based on the above graph you can see that the most important predictive variable is wheat. It separates the set of all records of the value 5.5. The most important decision rules of this analysis is as follows: if the variable V1 is less than 5.5, ie the area of wheat crops is less or equal to of 1,129,750 ha and the variable V3 - maize values is greater or equal to 280,337.5 ha, the lowest N_2O emissions will be obtained.

The other end of the graph indicates conditions of the highest possible emissions. They may take place in the increase of wheat crops over 2,414,180 ha, while the potato crops at less or equal than 803,384 ha and taking into account the secondary diversity of crops wheat - for areas over 2,476,870 ha.

In conclusion, the graph analysis suggests that wheat, maize and potatoes have an impact on the N_2O emissions value.

The total insights into CH_4 emission projections are presented below in the form of a decision tree (Figure 4). The most important variables affecting the emissions are V1-wheat, V3 – maize, V4 -sugar beet and V5 - potatoes.

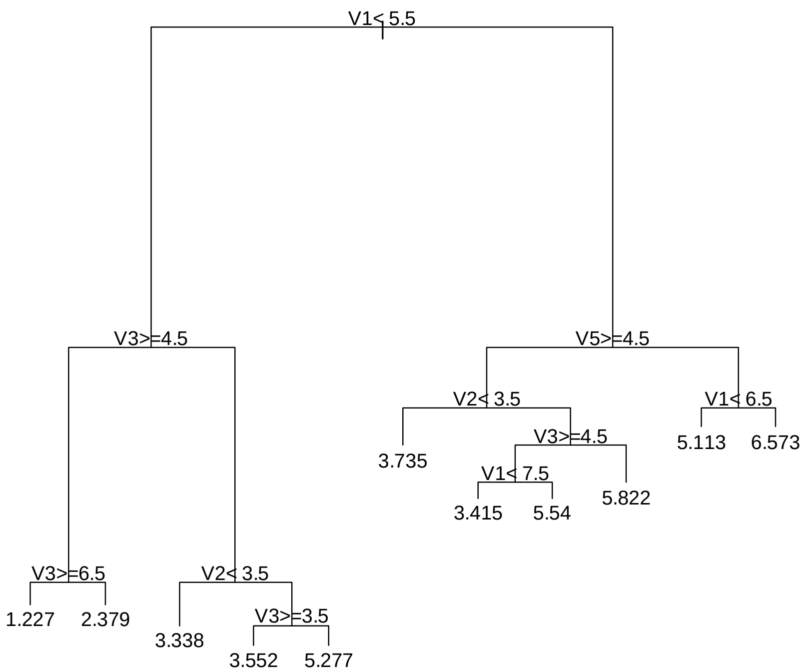

Fig. 3. Tree graph of N_2O emissions depending on the size of the of crops, where: V1 - wheat, V2 - rye, V3 - maize, V4 - sugar beet, V5 – potatoes; own calculations

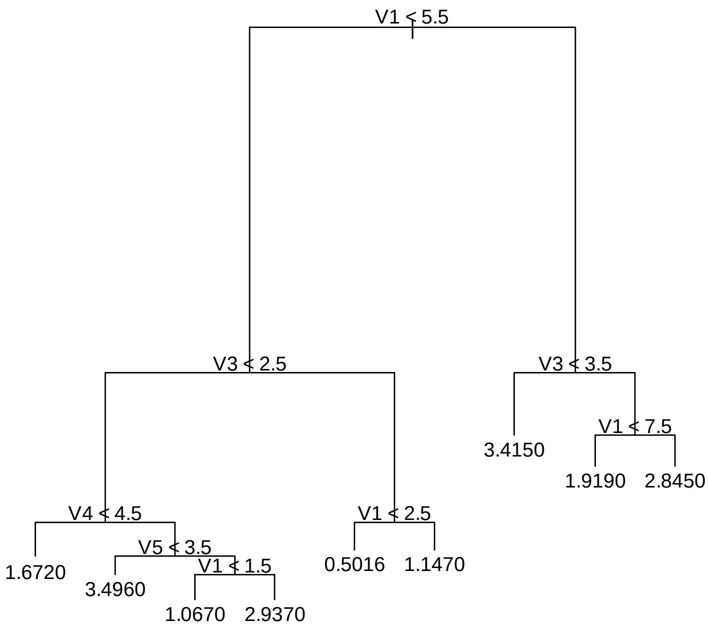

Fig . 4. Tree graph of CH_4 emissions depending on the size of the of crops, where: V1 - wheat, V2 - rye, V3 - maize, V4 - sugar beet, V5 – potatoes; own calculations.

In the present case the most important predictive variable is also wheat. As in the previous analysis, the set was divided of 5.5 value. The lowest CH_4 emissions were estimated when the variable V3 was greater than 2.5, ie maize crop is less or equal to 85,180 ha, and for V1 - wheat for less than 2,280,650 ha. In summary, the lowest emissions of CH_4 have been noted with the decline in wheat crops but simultaneously with the increase in maize crop.

Two cases were observed which in various conditions achieved the highest scale of CH_4 emissions. In the first instance, where the variable V1 (wheat) is less than 2,406,790 ha, V3 (maize) is less than 152,273 ha, V4 (sugar beet) more than 333,131 ha, and for V5 - potatoes less than 713,250 ha the highest emissions were noted. Emissions of very close values were observed also for the wheat over 2,414,180 ha and maize crops below 70,123 ha.

5. Conclusions

The paper presents an example of using Data Mining Tools to analyze the variables describing the amount of CH_4 and N_2O gases released from the agricultural sector, including arable crops areas. For this purpose predictive capabilities of neural network of Flexible Bayesian Models and modelling based on a decision tree using the *tree* function in an R-Project were used.

It has been shown that the analysis of data in a multidimensional approach extends the existing information in this field.

The decision trees devised within the framework of this analysis, provide the knowledge and methods that allow a rapid identification of the most important arable crops that affect the N_2O and CH_4 emissions released from the agricultural sector. The most important differentiating variables are wheat, maize and potatoes for N_2O emission and wheat and maize for CH_4 emission.

The noticeable difference in the two cases of classification tress is the scale of generated gases emissions - much more intensive, in a wide spectrum range for N_2O emissions (from 1.23 to 6.6) than CH_4.

6. References

[1] Bebkiewicz K., Cieślińska J., Dębski B., Jędrysiak P., Kanafa M., Kargulewicz I., Olecka A., Olendrzyński K., Rutkowski J., Sędziwa M., Skośkiewicz J., Sowińska K., Żaczek M.: Krajowy Raport Inwentaryzacyjny 2012, Inwentaryzacja gazów cieplarnianych dla lat 1988-2010, KOBiZE, Warszawa, 2012.

[2] Biecek P.: Na przełaj przez Data Mining; http://www.biecek.pl/R/naPrzelajPrzezDM.pdf, [accessed: 7.01.2013].

[3] Burczyk P., Miatkowski Z., Turbiak J.: Wstępne rozpoznanie emisji N_2O w wybranych siedliskach łąkowych w różnych regionach Polski, Woda-Środowisko-Obszary Wiejskie, 11, 4 (36): 57–67, 2011.

[4] De'ath G., Fabricius K.E.: Classification and regression trees: a powerful yet simple, Ecology, 81(11): 3178–3192, 2000.

[5] Feldesman M.R.: Classification Trees as an Alternative to Linear Discriminant Analysis, American Journal of Physical Anthropology, 2002, 119: 257–275.

[6] Geherke J., Rmakrishnan R., Ganti V.: RainForest - A Framework for Fast Decision Tree Construction of Large Datasets, Data Mining and Knowledge Discovery, Kluwer Academic Publishers, 2000, 4: 127–162.

[7] Gibbs M.J., Conneely D., Johnson D., Lasse K.R., Ulyatt M.J.: CH4 emissions from enteric fermentation, in IPCC, Background Papers: IPCC Expert Meetings on Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories. IPCC National Greenhouse Gas Inventories Programme, Technical Support Unit, 2002, 297–320.

[8] Hand, D.J.: Construction and Assessment of Classification Rules. Chichester, England: JohnWiley & Sons, 1997.

[9] Jun P., Gibbs M., Gaffney, K.: CH_4 and N_2O emissions from livestock manure in IPCC, Background Papers: IPCC Expert Meetings on Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories. IPCC National Greenhouse Gas Inventories Programme. Technical Support Unit, 2002, 321–338.

[10] Kajungu D. K., Selemani M., Masanja I., Baraka A., Njozi M., Khatib R., Dodoo A., Binka F., Macq J., D'Alessandro U., Speybroeck N.: Using classification tree modelling to investigate drug prescription practices at health facilities in rural Tanzania, Malaria Journal, 2012, 11:311.

[11] Kolasa-Więcek A.: Analiza zależności i siły związku między zużyciem nawozów azotowych a rodzajem upraw w Polsce, Journal of Research and Applications in Agricultural Engineering, 2011, 56(2): 87-89.

[12] Koronacki J., Ćwik J.: Statystyczne systemy uczące się, WNT, 2005.

[13] Kozłowski R.J, Weres J.: Komputerowe wspomaganie identyfikacji szkodników i chorób rzepaku ozimego, Inżynieria Rolnicza, 2007, 2(90): 109-117.

[14] Marcinkowski T.: Emisja gazowych związków azotu z Rolnictwa, Woda-Środowisko-Obszary Wiejskie, 2010, 10, 3 (31): 175–189.

[15] Monni S., Perälä P., Regina K.: Uncertainty in agricultural CH4 and N_2O emissions from Finland – possibilities to increse accurancy in emission estimates, Mitigation and Adaptation Strategies for Global Change, 2007, 12, 4: 545-571.

[16] Neal R.: Flexible Bayesian Models on Neural Networks, Gaussian Processes, and Mixtures v. 2004-11-10. University of Toronto, Toronto.

[17] Olendrzyński K, Kargulewicz I, Skośkiewicz J, Dębski B, Cieślińska J, Olecka A, Kanafa M, Kania K, Sałek P. Krajowa inwentaryzacja emisji i pochłaniania gazów cieplarnianych za rok 2007. KASHUE, 2009.

[18] Pasyniuk P.: Ograniczenie emisji gazów cieplarnianych przez zastąpienie oleju napędowego olejami roślinnymi, Problemy Inżynierii Rolniczej, 4: 79-89, 2010.

[19] Pietrzak S., Sapek A., Oenema O.: Ocena emisji podtlenku azotu (N_2O) ze źródeł rolniczych w Polsce. Zeszyty Edukacyjne 8. IMUZ, 2002, 23–36.

[20] Roszkowski A.: Technologie produkcji zwierzęcej a emisje gazów cieplarnianych, Problemy Inżynierii Rolniczej, 2: 83-97, 2011.

[21] Sapek A., 2008. Emisja tlenków azotu (NO_x) z gleb uprawnych i ekosystemów naturalnych do atmosfery. Woda Środowisko Obszary Wiejskie, 2008, 8, 1(22): 283–304.

[22] Smith K., Bouwman L., Braatz B.: Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories N_2O: Direct Emissions from Agricultural Soils, 361-380.

[23] 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, 2009, 14(4): 323-48.

[24] Velthof G.L, Oudendag D., Witzke H.P., Asman W.A.H., Klimont Z., Oenema O.: Integrated Assessment of Nitrogen Losses from Agriculture in EU-27 using MITERRA-EUROPE, Journal of Environmental Quality, 2009, 38, 2: 402-417.

[25] Wei-Yin L.: Improving the precision pf classification trees, The Annals of Applied Statistics, 2009, 3, 4: 1710–1737.

[26] Wei-Yin L., Yu-Shan S.: Split Selection Methods for Classifcation Trees, Statistica Sinica, 1997, 7: 815-840.

[27] Zaliwski A.: Emisja gazów cieplarnianych przez rolnictwo. Studia i Raporty, IUNG-PIB, 2007, 4: 35–46.

[28] Zimroz R., Brzychczy E.: Zastosowanie drzew decyzyjnych do wspomagania rozpoznawania stanu technicznego wybranych elementów maszyn górniczych, Bezpieczeństwo Pracy i Ochrona Środowiska w Górnictwie, 2010, 9: 30-33.

[29] FAO: http://faostat.fao.org/site/573/DesktopDefault.aspx?PageID=573#ancor, [accessed: 15.12.2012].

[30] UNFCCC: http://unfccc.int/di/DetailedByGas/Event.do?event=go, [accessed: 15.12.2012].