Cent. Eur. J. Chem. • 11(2) • 2013 • 259-270
DOI: 10.2478/s11532-012-0147-6

VERSITA

Central European Journal of **Chemistry**

C&RT model application in classification of biomass for energy production and environmental protection

Research Article

Marcin Sajdak*, Olaf Piotrowski#

*Institute for Chemical Processing of Coal,
41-803 Zabrze, Poland*

Received 8 August 2012; Accepted 3 October 2012

Abstract: Biomass is most often used to produce energy *via* its combustion and co-combustion along with conventional energy carriers. The prevalence of this method results from the lack of sufficient facilities that can provide a quick and simple chemical classification method, which would show a broader range of possible applications of biofuel in the energy industry. The aim of this study was the development of novel method of classification allowing for quick determine of the direction of the biomass usage by applying classification and regression models of trees (C&RTs). The proper functioning of a C&RT model is based on a very large database of results collected by the Institute for Chemical Processing of Coal during years of work in this field. The created model may be used as decision tool for grouping various biomass sources with respect to their further application in energy generation.

Keywords: *Biomass • Data Mining • Chemometric analysis • Energy • C&RT*

1. Introduction

Because of progressive global warming associated with CO_2 emissions, and the depletion of fossil fuels, researchers have been forced to seek innovative technologies to enable the use of environmentally friendly, renewable fuels. Biomass is an example of such an energy source. It is characterised by low environmental demand which implies straightforward availability. Moreover, the European Union is requiring the utilization of renewable energy sources, including biomass.

Recently, the most common method of utilizing biomass in the production of energy, has been its combustion and co-combustion along with conventional energy carriers. The use of combustion methods results from the lack of sufficient facilities that can provide a quick and simple chemical classification method to give a broader range of the potential applications of biofuels in the energy industry. Prolonged work by the Institute for Chemical Processing of Coal (ICHPW) has resulted in an innovative method to determine the possible applications of biomass through the use of an integrated model based on classification and regression trees (C&RTs) in conjunction with the use of linear estimation elements. To permit the proper operation of the C&RT model, the data were obtained from an information bank of physical and chemical biomass properties collected by ICHPW over a period of years. The created model may be used as a decision tool for grouping various biomass sources with respect to their further applications in energy generation. Despite other multiple purposes of the C&RT model, this article describes one pathway for biomass utilization through the production of bio-oils followed by the production of hydrocarbon-originated organic compounds.

2. Theoretical details

2.1. Classification Trees as a tool

The idea of classification trees relies on the gradual division of a set of arguments until uniformity is attained

* E-mail: msajdak@ichpw.zabrze.pl
# E-mail: opiotrowski@ichpw.zabrze.pl

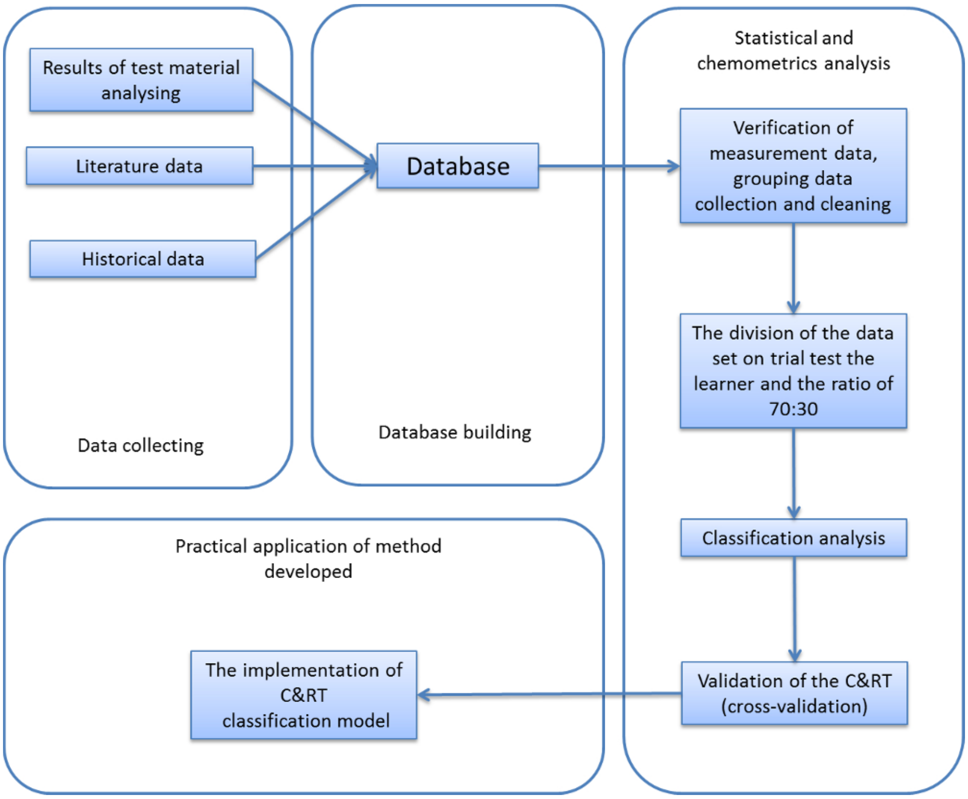

Figure 1. Example of a diagram to prepare the data and proceedings in the analysis.

in created subsets. The tree resembles a graph that consists of a root node from which at least two branches emerge, which then lead to inferior nodes (child nodes). Each node is attributed to a class description, and each branch refers to a decision rule, *i.e.*, a condition related to arguments from an entry data set and describing the case when each branch is chosen. Child nodes become parent nodes during successive splitting of the data set. Each division is performed for separate features (parameters). In a common algorithm, the conditions on the branches of each node must be complementary in a manner that provides one possible path downward when "climbing the tree". Nodes that do not have any child nodes are known as leaf, outer or terminal nodes, and represent the final classes.

Classification trees might be considered a collection of rules that enable separate sets of arguments to be linked into a common class. The path leading from the root node to the leaf node represents the conjugation of tests (complex). Using the tree to classify new objects relies on walking from one tip to another and down the branches that meet the features of the new item.

A diagram of the processes during the collection and analysis of measurement data has been presented in the Fig. 1. The first step was to collect data from different sources and then put them into a database. The database was constructed using a method which enables easy data archivisation and cooperation with statistical and chemometric software. The next step was to verify the data set and the rejection of erroneous data. Then aforementioned data were divided into two groups: one of them was a learn-group and the second one was a test-group. At the same time, results of the C&RT were cross-validated or verified using a predictive model on a test set. The implementation of the C&RT, and further testing of the new data sets was included in the final step.

Fig. 2 presents a simple example of a tree model used for the assignment of an energetic class of biomass samples under consideration. This classification does not require deep knowledge about every feature of the tested object, which significantly increases the practical applications of this classification method. This method might be most useful during the classification of biomass having a lack of a complete set of qualitative and quantitative data. The analysis of classification trees is one of the elemental techniques used in so-called data mining [1].

These techniques are used in classification-type problems. Classification-type problems are generally those where we attempt to predict values of a categorical

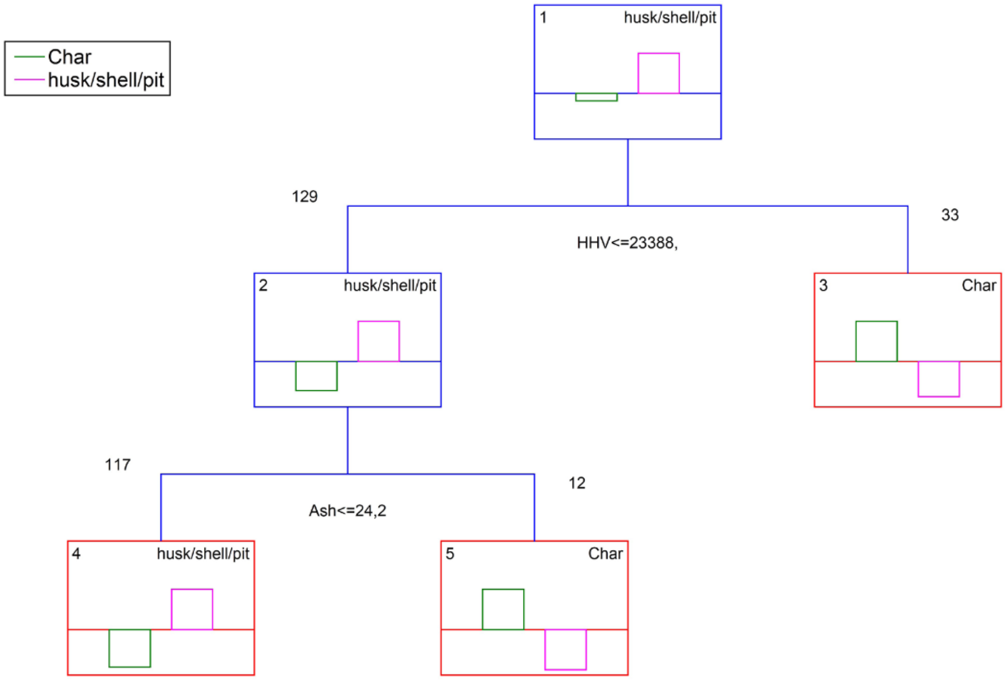

Figure 2. Example of a regression and decision tree.

dependent variable (class, group membership, *etc.*) from one or more continuous and/or categorical predictor variables. For example, we may be interested in predicting which sample is both from natural materials and which is characterized by zero CO_2 emission. These would be examples of simple binary classification problems, where the categorical dependent variable can only assume two distinct and mutually exclusive values. In other cases, we might be interested in predicting which one of multiple different alternative testing samples (biomass, bio-char, coal-char, hard coal) will be better either to generate energy or to produce liquid fuel. In those cases, there are multiple categories or classes for the categorical dependent variable.

Analyses based on C&RT aim to predict and explain the responses of a categorical dependent variable, which is why the tools used in this module share numerous common features with techniques from more conventional methods, such as discriminant analysis, cluster analysis, nonparametric statistics and nonlinear estimation.

3. Experimental procedure

3.1. Research methodology

A classification tree is created using recurrent divisions of the input data set into consecutive subsets until uniformity is attained. The main purpose for the creation of subdivisions is to obtain a collection of objects that are as uniform as possible with respect to a given dependent variable [2]. In the present work, a C&RT model was used to divide and classify biomass according to the best qualitative and quantitative characteristics that indicate the possibility of its use in the production of high-quality biofuel.

The general procedure of within the classification tree creation using the C&RT algorithm consists of a few steps:

1. Verification of uniformity between objects in matrix A via variant analysis or principal component analysis. If the matrix appears uniform at this stage, the work is finished; if not, proceed to the next steps.

2. Determination of possible partitioning of matrix A into homogeneous subsets B_1-B_n.

3. Qualitative analysis of each subset B_1-B_n according to established criteria and selection of the best subset.

4. Division of matrix A according to the chosen standard.

The last step of data-set division is based on the features that characterize the objects. These features must not be chosen randomly in order to avoid the situation where the number of leaf nodes is equal to the number of objects [3]. Hence, the selection of features that control the classification is constructed according to various statistical benchmarks or according to other information theory techniques. In this work, we applied variable priority analysis and Pareto diagram analysis,

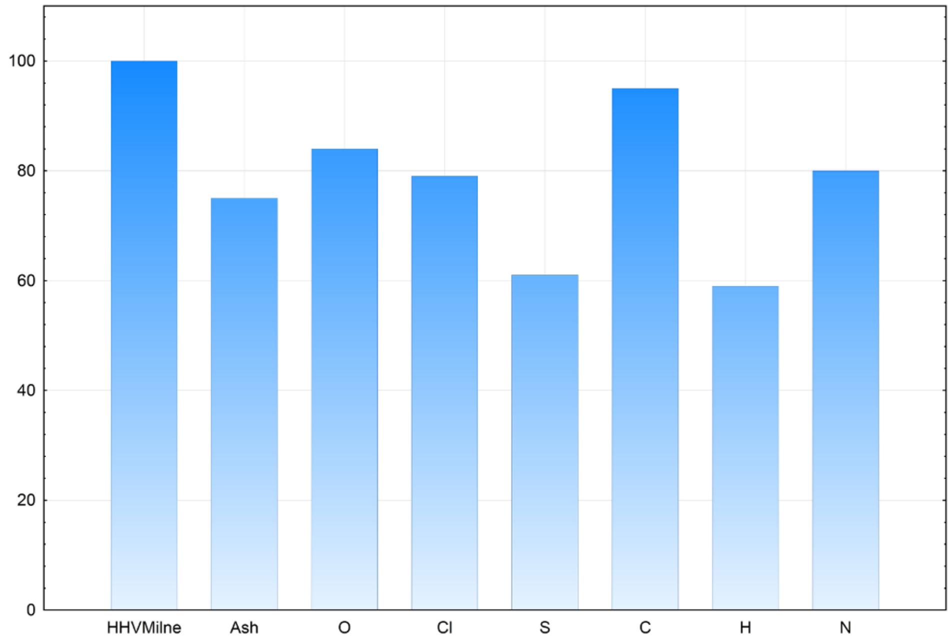

Figure 3. Example of a Pareto diagram of chosen variables.

which represent statistically significant variables that influence the qualitative parameter under consideration [4,5], as the method for feature selection.

Before the presumptive rejection of variables, the characterising features to be specified were:

- Quantity of features that would be used to divide the set into classes.

- Degree of division (size of the tree) necessary to obtain the tree with the smallest number of nodes without the loss of classification quality

- Method of object allocation from the root node to the subsets.

The application of these rules would lead to a tree characterized by the highest possible uniformity of objects in the created subsets and by the lowest number of nodes that lead to a set of simple classification rules.

3.2. Material, methods and research progress

The subject of research (under classification) is a database of biomass analysis results collected by ICHPW over a period of years combined with published data. The subject includes biomass, carbonised biomass, bio-oils and fossil fuels, such as bituminous coal, brown coal (lignite), peat and coke. The tested population comprises 1068 objects which are described by the following parameters:

- Ash content
- Enthalpy of combustion and calorific value
- Percentage content of:
- carbon
- hydrogen
- nitrogen
- oxygen
- sulphur
- chlorine

Example properties of study material analysis are presented in Table 1.

The classification procedure was intended to specify groups of biomass that are as homogeneous as possible, taking into account crucial features related to later processing by pyrolysis in the production of bio-oils. Except using of biomass as a raw material for pyrolysis and also in producing bio-oil, this material is more and more frequently used in the gasification process [6-8].

Application of biomass for energy production is ecologically and economically justified due to the zero CO_2 emission. The use of biomass reduces the amount of carbon dioxide emitted to the atmosphere during combustion and thus reduced emission costs [9].

The first stage of research required the selection of appropriate diagnostic parameters essential for the classification of biomass for pyrolysis. The choice of these parameters considered features such as the aim of classification, the process characteristics, the statistic rating of variables, data integrity and analysis of variable importance *via* the Pareto method (Fig. 4). The following variables were chosen to establish the suitability of the biomass for the production of bio-oils:

- Heat of combustion – HHV,
- Ash content – Ash,

Table 1. Results of study material analysing.

Group	Ash	HHV	Energy group MJ kg^{-1}	C	H	O	O/C	N	S	Cl
Char	45.90	10621.00	10-15	28.20	1.81	8.30	0.29	0.24	0.02	9.08
Char	67.10	12472.00	10-15	33.50	1.32	2.20	0.07	1.01	0.56	2.92
Char	37.90	17412.00	15-20	46.00	2.80	12.50	0.27	1.20	0.41	2.00
Char	49.60	15294.00	15-20	42.50	1.40	5.00	0.12	0.80	0.49	2.64
Char	20.90	20082.00	20-25	56.40	2.58	18.90	0.34	0.66	0.55	0.00
Char	7.40	23734.00	20-25	64.60	3.40	21.00	0.33	1.50	0.14	0.07
Char	7.70	28352.00	25-30	77.60	2.50	9.50	0.12	1.50	0.20	0.27
Char	4.10	26071.00	25-30	68.40	4.10	21.30	0.31	0.50	0.04	0.03
Char	6.20	29866.00	25-30	81.70	2.40	8.60	0.11	0.40	0.04	0.05
Coal	20.40	25356.00	25-30	65.30	3.53	10.00	0.15	1.62	0.92	0.01
Coal	11.00	28323.00	25-30	69.00	5.00	12.80	0.19	1.50	0.67	0.05
Coal	13.70	27644.00	25-30	70.00	4.00	10.00	0.14	1.50	0.73	0.03
Coal	2.90	34370.00	od30	88.90	3.40	2.30	0.03	1.55	0.81	0.08
Coal	6.20	30872.00	od30	76.70	4.69	10.50	0.14	1.41	0.40	0.06
Coal	5.10	32080.00	od30	79.20	4.70	7.60	0.10	1.80	0.90	0.70
Coal	12.10	29691.00	25-30	71.90	4.90	8.60	0.12	1.50	1.00	0.01
Coal	12.80	27823.00	25-30	69.30	4.30	10.70	0.15	1.20	1.48	0.24
Coal	13.20	25807.00	25-30	64.60	4.20	13.70	0.21	1.30	2.90	0.12
Coal	7.40	28294.00	25-30	70.80	4.70	15.90	0.22	0.70	0.50	0.00
Coal	8.30	30604.00	od30	75.50	4.70	9.40	0.12	1.30	0.70	0.08
Peat	7.50	21747.00	20-25	53.50	5.80	32.00	0.60	2.00	0.32	0.05
Peat	6.30	21350.00	20-25	53.30	5.60	33.20	0.62	1.43	0.16	0.06
Peat	2.70	19983.00	15-20	51.20	5.60	39.50	0.77	0.90	0.10	0.02
Peat	4.30	21694.00	20-25	54.50	5.60	33.60	0.62	1.80	0.20	0.03
Peat	3.80	21876.00	20-25	55.20	5.70	35.50	0.64	1.50	0.19	0.04
Grass/plant	10.30	17147.00	15-20	45.10	4.97	35.60	0.79	3.30	0.16	0.56
Grass/plant	10.70	17657.00	15-20	45.50	5.15	34.50	0.76	3.29	0.30	0.54
Grass/plant	6.90	17777.00	15-20	46.30	5.29	39.30	0.85	1.73	0.12	0.30
Grass/plant	10.10	17412.00	15-20	45.20	5.14	36.10	0.80	2.91	0.23	0.31
Grass/plant	9.20	18170.00	15-20	46.30	5.37	35.70	0.77	2.89	0.26	0.31
Grass/plant	10.70	17904.00	15-20	46.20	5.12	35.30	0.76	2.08	0.14	0.48
Grass/plant	9.40	18556.00	15-20	45.00	6.00	36.90	0.82	2.50	2.00	0.30
Grass/plant	7.30	17976.00	15-20	46.80	5.40	40.70	0.87	1.00	0.02	0.03
Grass/plant	5.30	19047.00	15-20	47.20	6.00	38.20	0.81	38.20	2.68	0.50
Grass/plant	5.20	18800.00	15-20	46.80	5.95	38.50	0.82	2.88	0.19	0.50
Husk/shell/pit	6.20	18712.00	15-20	46.50	5.97	40.10	0.86	1.15	0.04	0.05
Husk/shell/pit	5.80	17630.00	15-20	45.80	5.36	40.60	0.89	0.96	0.01	0.08
Husk/shell/pit	3.00	20015.00	20-25	50.10	5.95	40.10	0.80	0.74	0.03	0.01

Continued **Table 1.** Results of study material analysing.

Group	Ash	HHV	Energy group MJ kg^{-1}	C	H	O	O/C	N	S	Cl
Husk/shell/pit	0.90	20097.00	20-25	49.70	6.30	42.80	0.86	0.26	0.01	0.00
Husk/shell/pit	6.20	17965.00	15-20	46.50	5.45	41.00	0.88	0.71	0.03	0.05
Husk/shell/pit	5.40	18754.00	15-20	46.70	5.98	40.80	0.87	1.01	0.05	0.03
Husk/shell/pit	6.10	19181.00	15-20	47.50	5.97	39.20	0.83	1.13	0.06	0.02
Husk/shell/pit	6.20	18711.00	15-20	46.50	5.97	40.10	0.86	1.15	0.04	0.05
Husk/shell/pit	4.90	18994.00	15-20	47.70	5.87	40.00	0.84	1.45	0.07	0.07
Husk/shell/pit	2.80	17956.00	15-20	46.20	5.79	44.50	0.96	0.65	0.04	0.02
Plastics	0.40	20334.00	20-25	41.20	5.28	5.80	0.14	0.04	0.03	47.70
Nonorganic residue	16.20	18877.00	15-20	45.10	5.78	30.10	0.67	2.77	0.12	0.41
Nonorganic residue	60.80	9686.00	0-10	23.40	2.65	10.70	0.46	1.49	0.57	0.41
Nonorganic residue	56.20	11883.00	10-15	28.40	3.01	10.70	0.38	0.94	0.03	0.81
Nonorganic residue	11.20	26527.00	25-30	56.80	7.47	18.00	0.32	3.44	0.16	2.92
Nonorganic residue	12.60	31165.00	od30	67.70	6.76	5.90	0.09	0.35	0.74	5.88
Nonorganic residue	35.00	18647.00	15-20	41.60	5.11	14.20	0.34	2.15	0.12	1.79
Nonorganic residue	42.50	18093.00	15-20	38.30	5.32	11.70	0.31	2.02	0.32	0.56
Nonorganic residue	10.60	30616.00	od30	67.40	6.84	7.70	0.11	2.94	0.11	3.14
Nonorganic residue	18.90	25639.00	25-30	56.60	5.97	8.20	0.14	2.81	0.10	5.64
Nonorganic residue	1.20	37298.00	od30	82.20	7.21	1.10	0.01	0.99	0.03	1.90
Nonorganic residue	1.20	36680.00	od30	81.50	7.09	3.20	0.04	0.66	0.02	0.69
Untreated wood	3.30	19722.00	15-20	50.40	5.64	40.10	0.80	0.55	0.03	0.02
Untreated wood	3.00	20756.00	20-25	51.60	6.00	39.30	0.76	0.10	0.02	0.00
Untreated wood	4.00	20055.00	20-25	50.30	5.83	39.60	0.79	0.11	0.07	0.03
Untreated wood	8.50	18899.00	15-20	47.10	5.74	38.10	0.81	0.39	0.11	0.01
Untreated wood	9.40	18297.00	15-20	47.30	5.20	37.70	0.80	0.40	0.05	0.03
Untreated wood	10.60	17789.00	15-20	46.30	5.07	37.50	0.81	0.47	0.08	0.02
Untreated wood	11.20	18329.00	15-20	46.30	5.40	36.60	0.79	0.47	0.02	0.02
Untreated wood	5.10	17700.00	15-20	48.80	4.60	41.00	0.84	0.33	0.08	0.04
Untreated wood	5.30	19278.00	15-20	49.70	5.40	39.30	0.79	0.20	0.10	0.01
Treated wood	5.90	19159.00	15-20	49.80	5.24	38.60	0.78	0.37	0.03	0.01
Treated wood	20.10	14657.00	10-15	40.20	4.10	34.60	0.86	0.69	0.12	0.17
Treated wood	42.80	10871.00	10-15	29.00	3.26	23.80	0.82	0.89	0.15	0.04
Treated wood	33.00	12640.00	10-15	34.00	4.20	33.80	0.99	1.02	0.15	0.35
Treated wood	2.00	19321.00	15-20	48.40	5.93	40.60	0.84	1.06	0.06	0.05

Continued **Table 1.** Results of study material analysing.

Group	Ash	HHV	Energy group MJ kg^{-1}	C	H	O	O/C	N	S	Cl
Treated wood	9.40	17898.00	15-20	45.50	5.51	37.80	0.83	1.79	0.03	0.01
Treated wood	4.00	22989.00	20-25	47.90	8.59	37.50	0.78	1.12	0.01	0.98
Treated wood	2.80	19051.00	15-20	48.10	5.93	42.60	0.89	0.47	0.11	0.02
Treated wood	2.10	19070.00	15-20	49.90	5.43	42.00	0.84	0.45	0.04	0.07
Treated wood	2.90	18909.00	15-20	48.70	5.64	40.20	0.83	2.35	0.08	0.16
Straw (stalk/ cob/ear)	4.90	18117.00	15-20	46.80	5.53	41.90	0.90	0.41	0.06	0.41
Straw (stalk/ cob/ear)	5.90	18021.00	15-20	46.90	5.31	40.10	0.86	0.73	0.12	0.98
Straw (stalk/ cob/ear)	4.30	18205.00	15-20	45.90	5.92	43.00	0.94	0.43	0.20	0.35
Straw (stalk/ cob/ear)	6.40	18592.00	15-20	46.10	5.93	40.10	0.87	0.78	0.33	0.39
Straw (stalk/ cob/ear)	2.70	18975.00	15-20	47.20	6.14	42.70	0.90	0.80	0.15	0.30
Straw (stalk/ cob/ear)	5.90	18186.00	15-20	46.20	5.70	41.30	0.89	0.60	0.08	0.27
Straw (stalk/ cob/ear)	1.40	18112.00	15-20	46.60	5.87	45.50	0.98	0.47	0.01	0.21
Straw (stalk/ cob/ear)	2.50	17359.00	15-20	43.40	6.17	45.80	1.06	1.02	0.93	0.13
Straw (stalk/ cob/ear)	5.10	18442.00	15-20	46.80	5.74	41.40	0.88	0.66	0.11	0.27
Straw (stalk/ cob/ear)	5.60	16885.00	15-20	43.70	5.56	43.30	0.99	0.61	0.01	0.60

- Oxygen content – O,
- Oxygen to carbon ratio – O/C.
- Chlorine content
- Sulphur content
- Nitrogen content
- Origin of the biomass

An unsymmetrical distribution of these data, which show significant deviation from normal data sets, is evident. Moreover, the variances reach high values and are accompanied by outlying values. Because the analyses were to be performed on processes with parameters characterised by unsymmetrical distributions, the C&RT calculation model was chosen instead.

Based on the previously mentioned analysis, the HHV parameter was described using six categories:

- Up to 10 MJ kg^{-1} s.m.
- Between 10 and 15 MJ kg^{-1} s.m.
- Between 15 and 20 MJ kg^{-1} s.m.
- Between 20 and 25 MJ kg^{-1} s.m.
- Between 25 and 30 MJ kg^{-1} s.m.
- Greater than 30 MJ kg^{-1} s.m.

This feature is the most significant and crucial variable because it determines the caloric value of the bio-oil produced *via* pyrolysis. The ash content in biomass was classified as one of two groups: up to 5% s.m. and greater than 5% s.m. When this parameter was used for classification, the group with the lowest amount of inorganic ballast, which is later encountered in carbonised material, was created.

The oxygen content and oxygen-to-carbon ratio, O/C, are significant parameters because of the characteristics of the pyrolysis process. Greater oxygen content (and greater O/C values) result in more water remaining after the reaction of the product (bio-oil). Such corrosive conditions can damage the apparatus, as well as requiring additional processing.

The last feature under consideration is the biomass origin type. This feature may represent from several to more than a dozen levels of values in the population: X_1-X_n. Its significance shows up when localization and other economic points of view are revised. Depending on the occurrence of biomass in a given region, the costs of purchasing and importing biomass might be classified as a substrate for pyrolysis.

After the parameters have been properly verified, the next step concerns the classification of the tested biomass via the construction of a proper tree with

Table 2. Results of statistical analysis of random parameters chosen for the tested data population. Source: Own compilation.

	Mean	Minimum	Maximum	CV	Std dev.	Skew
Ash	11.791	0.400	67.100	14.071	119.341	1.517
HHV	20832.651	9686.000	37298.000	5648.444	27.113	609.087
C	51.983	23.400	88.900	13.132	25.263	1.416
H	5.126	1.320	8.590	1.312	25.602	0.142
O	28.438	1.100	45.800	14.451	50.816	1.558
N	1.641	0.040	38.200	4.074	248.222	0.439
Cl	1.170	0.002	47.700	5.267	450.134	0.568
S	0.309	0.010	2.900	0.516	166.829	0.056
O/C	0.598	0.013	1.055	0.324	54.187	0.035

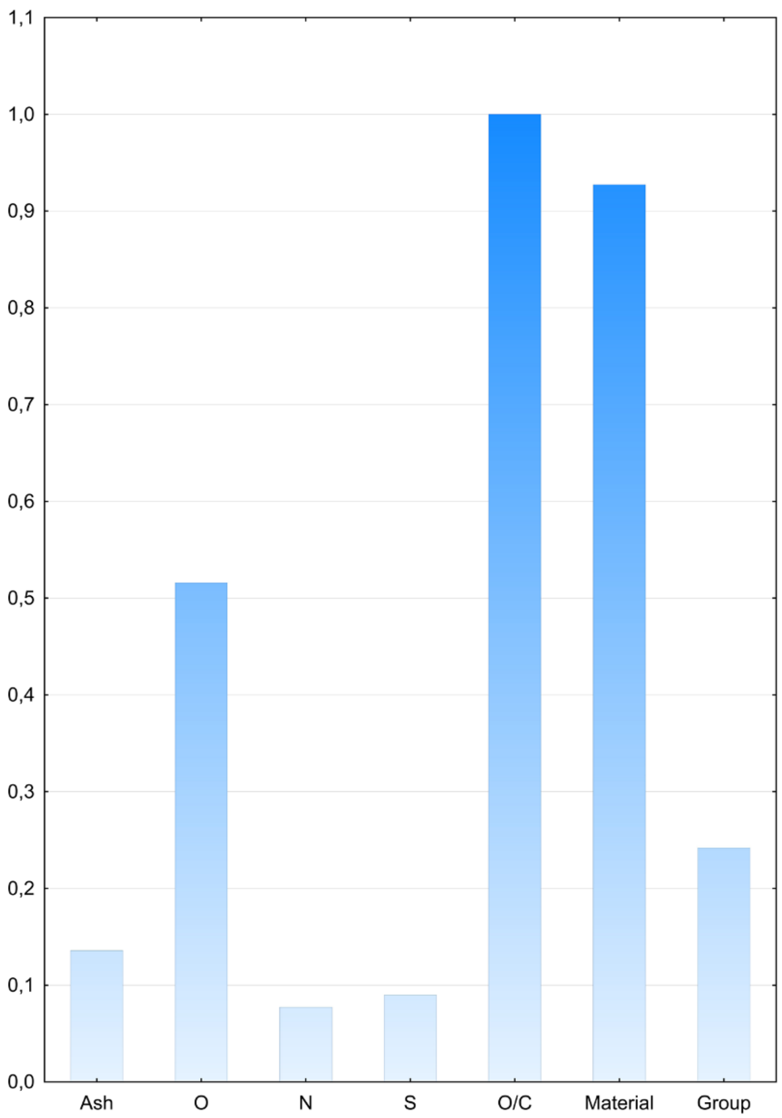

Figure 4. Pareto diagram of variables examined in research for process biomass classification.

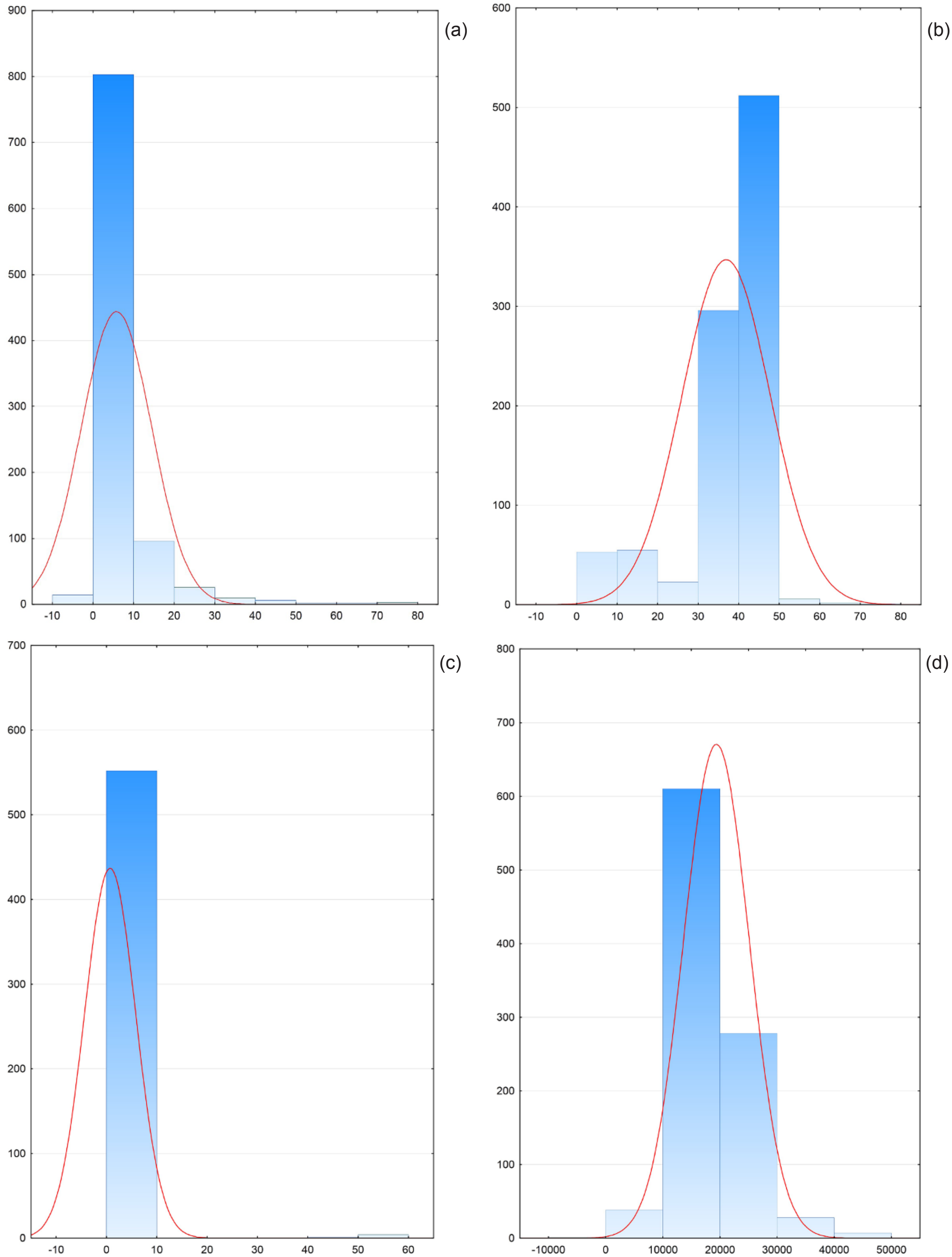

Figure 5. Histograms of selected variables: a) ash content, b) oxygen content, c) chlorine content, d) heat of combustion.

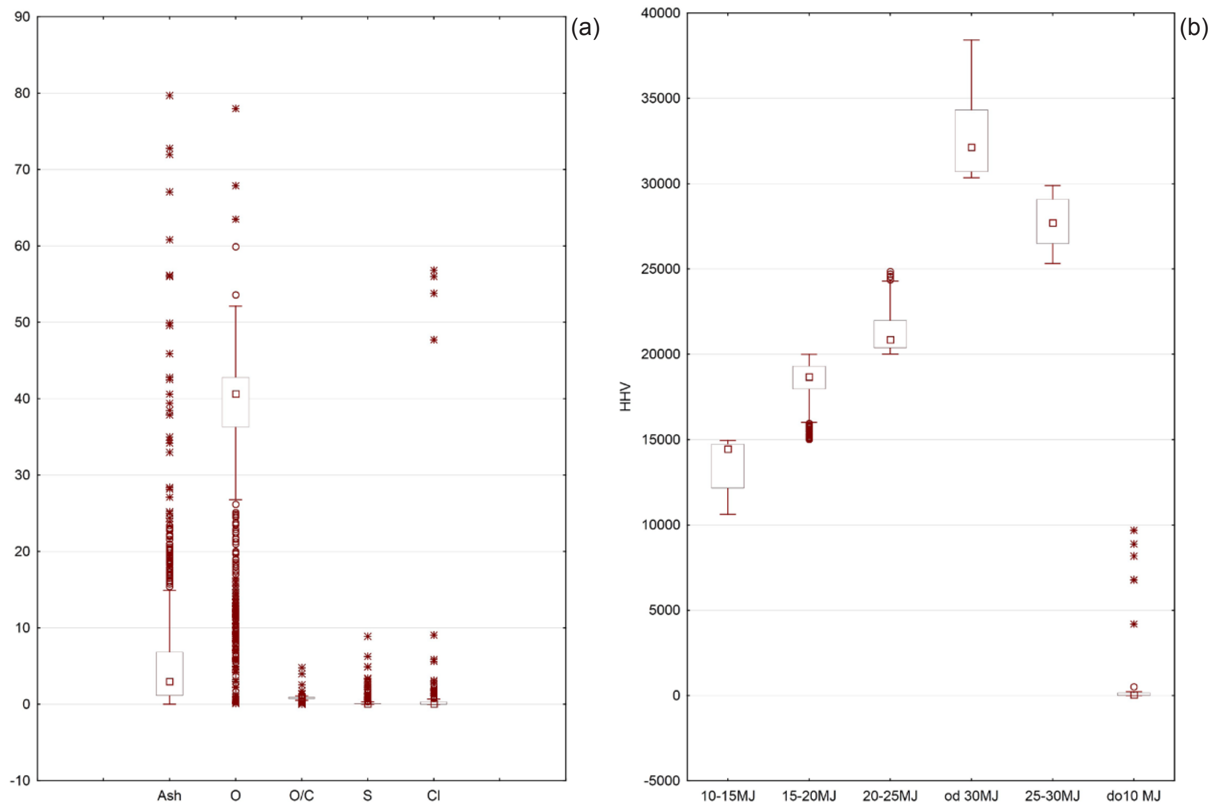

Figure 6. Exemplary charts of dispersion of variables chosen for analysis.

automatic C&RT algorithm procedures; this process thus devaluates the search for 1D assignation. During the creation of the tree, over-determination, *i.e.*, excessively exact matching to the input objects, may be encountered. Such a tree becomes too complex and does not provide sufficient generalization. Therefore, when the decision-taking space is created, the tree is cross-validated and properly stopped. The expansion of the tree is halted when the error of the validating group reaches a minimum.

Through the utilization of the process of building and arranging the classification tree, the highest possible uniform biomass subsets were obtained with respect to the chosen variables. The created diagram enables the simple construction of a set of rules in the form of an intersection of several logic conditions.

4. Results and discussion

Because of the application of the proper methods and chemometric procedures combined with data mining, a uniform classification tree was created that describes the algorithm for the classification of unknown biomass samples. A statistical description of these input and output variables is presented in Table 2. Because Table 2 does not provide unequivocal conclusions concerning the distribution of particular parameters for clarification, histograms of the selected variables are presented in Fig. 4. Fig. 6 presents 1-D charts of several variables, demonstrating the data dissipation and deviation from the mean values. The presented diagram is of the typical binary type, i.e., each parent node has up to two child nodes. The diagram consists of one root node and ten child nodes (leaves), as presented in Fig. 6. The root node, which contains 1060 objects, divided the population into 2 child nodes based upon the oxygen content. The left node (No. 2) represents subsets of highly energetic objects (bituminous coal, lignite, coke, peat, polymers – 102 objects). The right node (No. 3) contains mainly biomass – 840 objects. The next step split node No. 3 by the ash content into sets No. 4 (567 objects) and No. 5 (257 objects). Thus, subgroups of the various energetic characteristics were differentiated. Group No. 4, which exhibits greater values of heat of combustion, was divided into groups No. 6 and No. 7 based on the O/C ratio in the subsequent step. Examples of biomass with potential applications for bio-oil production are contained in group No. 6. As evidence for this statement, the histogram of the isolated population

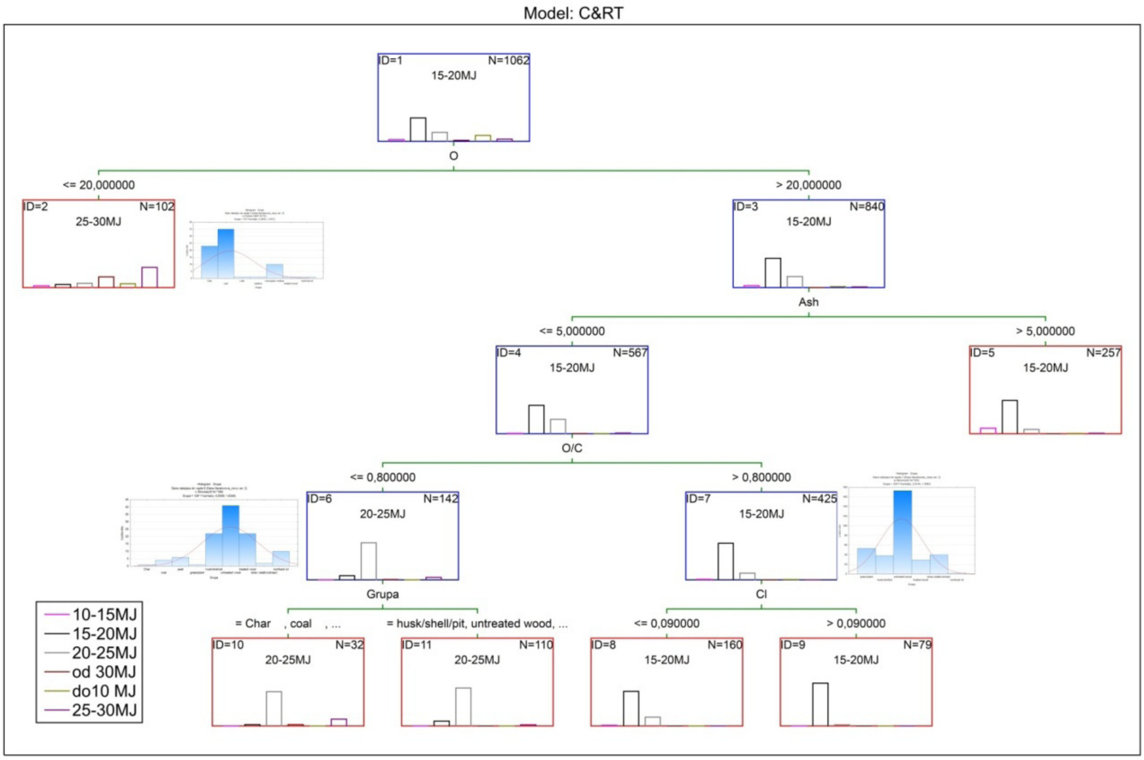

Figure 7. Final form of the decision-taking tree for the analysed trials,

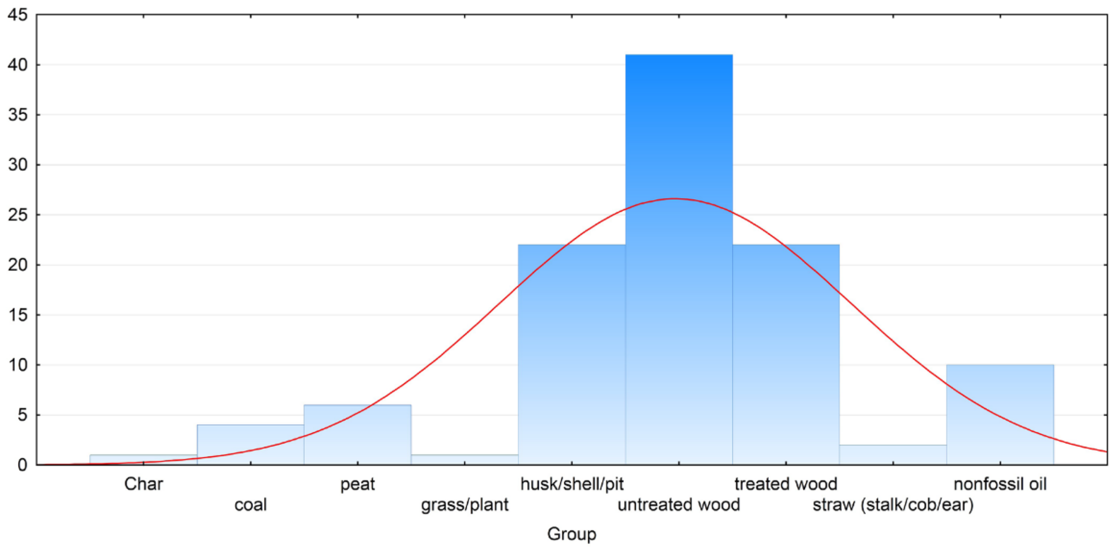

Figure 8. Histogram that represents the population differentiated in group No. 6 (child node).

is presented in Fig. 8, which includes the mentioned bio-oils (nonfossil oil) among other biomass.

Based on the classification tree scheme, 110 biomass examples among 1062 tested objects (more than 10% of all the samples collected into three groups) are suitable for direct application in the pyrolysis process. This group is characterized by apparently high values of heat of combustion, which range from 20 to 25 MJ kg^{-1}, and also by low ash content. The second notable group contains 160 objects (15%) with low chlorine content and heats of combustion between 15 and 20 MJ kg^{-1}. Biomasses from this class are suitable for pyrolytic conversion to bio-oils of inferior quality or for thermal conversion in a torrefaction process which decreases the volume of the biomass for bulk storage and increases the biomass energy density.

Table 3. The following table below presents the qualitative and quantitative analysis results.

	Observed	Predicted 15-20 MJ	Predicted 20-25 MJ	Predicted 25-30 MJ	Sum in line
Amount	15-20 MJ	339	10	4	353
Percent in collumn		82.89%	16.95%	10.26%	
Percent in line		96.03%	2.83%	1.13%	
Overall percent		66.86%	1.97%	0.79%	69.63%
Amount	20-25 MJ	47	46	2	95
Percent in collumn		11.49%	77.97%	5.13%	
Percent in line		49.47%	48.42%	2.11%	
Overall percent		9.27%	9.07%	0.39%	18.74%
Amount	25-30 MJ		2	19	21
Percent in collumn		0.00%	3.39%	48.72%	
Percent in line		0.00%	9.52%	90.48%	
Overall percent		0.00%	0.39%	3.75%	4.14%
Amount	Ogół grup	409	59	39	507
Percent sum		80.67%	11.64%	7.69%	

5. Conclusion

Based on the performed tests, the applied data mining, in conjunction with the C&RT method, provided satisfactory tools for the differentiation of biomass, which represents a complex data group. The characteristic feature of this methodology is its simultaneous division of the group of objects into classes and the establishment of these groups via simple rules of adherence.

The complete statistic description of the formed groups was obtained, which significantly helps in object classification during the deduction process. With this method, the quick classification of new objects is achieved, which is particularly advantageous when the operator lacks detailed knowledge of the material being classified.

The use of the C&RT model as a hierarchical tool for differentiation of multi-attribute objects confirmed that the construction of this type of model can be easily applied to biomass classification in which significant criteria related to a sustainable biomass economy are considered. In the examined biomass samples, the dominant parameter was the heat of combustion, which was approximately 15-20 MJ kg^{-1} d.m. This parameter determines the direction of further processing. With this method of categorization, the composition of biomass type, as related to different processing options, can be easily assigned, which could remarkably increase the efficiency of a processes such as pyrolysis. By linking the C&RT with rules of regression-model formulation, a predictive model of bio-oil parameters was determined as a function of variables that describe the biomass classified using the C&RT method.

References

[1] R. Lehtonen, E. Pahkinen, Practical methods for design and analysis of complex surveys (Wiley & Sons, England, 2004)

[2] R. Nisbet, J.F. Elder, G. Miner, Handbook of statistical analysis and data mining applications (Academic Press, Amsterdam, 2009)

[3] R.G. Brereton, Book-Applied chemometrics for scientists (Wiley & Sons, Chichester, 2007)

[4] G. Tao, P. Geladi, T.A. Lestander, S. Xiong, Renewable and Sustainable Energy Reviews 16(5), 3481 (2012)

[5] G. Tao, P. Geladi, T.A. Lestander, S. Xiong, Renewable and Sustainable Energy Reviews 16(5), 3507 (2012)

[6] A. Molino, G. Giordano, V. Motola, G. Fiorenza, F. Nanna, G. Braccio, Fuel (In Press), DOI: 10.1016/j.fuel.2012.06.104

[7] A.V. Bridgwater, Pyrolysis and Gasification of Biomass and Waste: Proceedings of an Expert Meeting: Strasbourg, France, 30 September - 1 October 2002, (Cpl Press, France, 2003)

[8] D.L. Giltrap, R. McKibbin, G.R.G. Barnes, Solar Energy 74, 85 (2003)

[9] C. Di Blasi, C. Branca, J. Anal. Appl. Pyrolysis 67, 207 (2003)