Modeling of Health State Quality in Regions

Jiří Křupka, Miloslava Kašparová, Jan Mandys, and Pavel Jirava

Abstract**—**The paper focuses on the problem of application of system approach on the quality of life modeling. Suggested quality of health state model works with real yearly data from 1997 to 2007 from the regions of the Czech Republic. Selected data from health, environmental and economic areas is used. Selected algorithms of cluster analysis and classification decision trees are applied by models creation.

Keywords**—**Cluster analysis, decision tree, environment, health, quality of life.

I. Introduction

We have to use system engineering, system approach [1], [2], [3] for a solution of comprehensive, complicated, and complex systems. A problem of modeling of the quality of life (QL) can be specified like this type of system [4], [5], [6] [7].

We can talk about QL in both the static and the dynamic meanings, the static meaning provides reports on peoples' life as of a set time, while the dynamic meaning compares and evaluates the QL in a longer period of time [8]. Generally it is possible to say that the QL is influenced by the physical and mental health of an individual, by the level of independence, by social attitude towards the environment and by other factors [8]. It may be defined as an individual's life satisfaction with the life the individual person lives compared to an ideal life. The evaluation of the QL depends in this case on the value system of each individual [8]. If we needed to make a complicated inquiry with each individual about the individual aspects of his/her life, such approach would be very time consuming and it would not be possible to execute research into the QL on an unlimited number of respondents. For this reason we are looking for methods that would allow us to draw data from publicly available databases and reach relevant results. There a universal delimitation of the QL does not exist. The QL depends also on external factors [9], good living conditions or other conditions define a high QL, but if such conditions vary, the satisfaction with the QL then also varies.

The QL [10], [11] is influenced by: material state of affairs (goods, services, home, economic level, by conditions for work and recreation, average income, purchasing power, and so); by the quality of the environment (the level of the utilization of natural resources, sustainable development [7], water quality, weather, soil, and other); by the individual's quality of health (the health of the society); by the quality of education, by moral and psychological climate (inside the family, organizations, culture, states), individual feeling of safety (physical, legal, societal) and by the possibility for self-expression.

According to [12] QL is defined as satisfaction of a concrete individual with achieving goals defining the orientation of his/her life. This orientation is further influenced by each individual's value orientation (as a hierarchy of values in the spiritual meaning).

Models designed in this work use data received from public sources. These are real data coming from the Czech Statistical Office, the Institute of Health Information and Statistics of the Czech Republic (CR) etc.

In this work we focus on the specification of Quality of Health State (QHS) in the CR regions, and modeling this "quality" on the basis of selected environment and economic attributes. In other words we deal with question "What the selected environment and economic attributes influence the health state (situation) in the CR regions?".

Objectives of this paper are:

- Create three linguistic levels of QHS in the regions based on real region input health indicators
- Modeling of QHS with selected environmental and economic indicators.

II. Problem Formulation

The term QL is discussed in many fields of science and each field of science approaches the term mainly from the view of its own science terminology. The value system of each individual has a major role in the definition of the individual's QL.

"Values represent a system of acquired dispositions of an individual to act towards or to strive to move towards a goal in accord with the desires determining the conditions of the existence." [13], p. 385. According to [14] values represent an individual's life purpose, they are a means for the adaptation to society and they are a means for conflict resolution inside a personality's system (that is in particular a conflict born in decision making). Values influence behavior, they integrate a personality in the context of important life goals and prevent chaotic behavior. At the same time they are also one of the sources of conflicts.

Schwarz's value system can be taken as the generally

This work was supported by the scientific research project of Ministry of Environment, CR under Grant No. SP/4i2/60/07 with title Indicators for Valuation and Modelling of Interactions among Environment, Economics and Social Relations, and it was supported by the National Science Foundation of CR under Grant No. 402/08/0849 with title Model of Sustainable Regional Development Management.

accepted value model. This model describes ten types of values. Schwarz has derived these value types by analyzing individual social needs. He considers these values to be essential for the functioning and the survival of all systems [15]. The model is based on the assumption that each individual is influenced by the following basic needs [16]:

- The necessity to fulfill his/her biological needs
- Participation in social interactions (coordinated interactions)
- To meet institutionalized requirements that are related to group survival possibilities.

Schwarz's model can be interpreted as a circle structure that consists of four areas [16]:

- Transcendence: it includes the values of universalism and benevolence
- Conservatism: it includes the values of conformity and related traditions and the value safety
- Orientation on self of an individual: it includes power, success and partially the value of hedonism, however, this value is included also in the following area
- Openness to change: it includes the already partially mentioned value of hedonism, but also the value of stimulation and self-determination.

It is essential to realize, in particular, the importance of cultural and social factors for the preservation of values that we recognize. The culture norms of a given society system give us clear guidance to what rules we must obey. It shows us the values and norms accepted by the majority society and it applies to all members of the given society [13].

Health represents a component of the overall QL that is generally valued and accepted in all societies. In [17] the author, on a general level, defines health as an abstract representing intact body and well functioning and good condition of all body organs and the entire body. Health is thus a normal function and illness means abnormal function, or suppressed function or an atypical function. Further, the author brings attention to the meaning of the definition of health as it is represented by the World Health Organization. The organization understands health to be the state of absolute body, mind and social well-being. Health understood in this way is not anymore the solely biological-medical problem, but is overlaps to social sciences.

The multidimensional notion of health is clearly demonstrated in [12]. It is stated there that physicians understand the meaning of health to be the absence of any illness or injury. Sociologists see the notion of health as the ability to function well in all social roles. Idealists describe a healthy individual as an individual that feels well physically, mentally, spiritually and socially. Humanists consider an individual to be healthy when the individual bears positively the burden of all daily life requirements and tasks to be fulfilled by the individual.

There are various ways how to measure the QL. Generally these methods of measurement can be divided into three groups [6] where:

- The evaluator is a third person (the shortcoming is that this evaluation is in many cases different from the evaluation of the individual done by himself/herself, one of the methods is Karnof index – e.g. a physician expresses his/her opinion on the total health status of the patient as of a certain date)
- The individual himself/herself is the evaluator – the QL is measured as it is subjectively felt and defined by the person who is himself/herself the subject of the inquiry. Here various methods are used, e.g. the evaluation of the individual quality method program (Schedule for the Evaluation of Individual QL) – the respondent fills into a questionnaire his/her personal life goals, evaluates the fulfillment of these goals and defines their importance, another method is the health method related to the QL (Health Related QL)
- The evaluation arrived at by the combination of the above-mentioned methods is used – here the method of the short way of assessment of the QL can be utilized (Manchester's Short Assessment of QL) - with this method we evaluate not only the complete life satisfaction, but also the satisfaction with a number of pre-defined life dimensions, or the Life Satisfaction Scale method can be used.

We can define QL system S_{QL} for design of QL model based on system approach by the following way:

$$S_{QL} = \{A_i, M_j, I_{ki}, Ap_m, De_n\} \quad (1)$$

where: A_i is i-th approach for description (specification) of QL, M_j is j-th method for QL modeling, I_{ki} je k-th indicator (attribute) for i-th approach, Ap_m je m-th appendix attribute and De_n je n-th demographic attribute.

If we focus on

- Economy approach, where the economic situation of an individual (the society) has more and more influence on his/her QL. The economic situation of a certain society can be expressed, among other indicators, by the well known gross domestic product (GDP) indicator. The higher the GDP indicator level, the better economic situation in a given country. Some critics of the expansive economy and of an unlimited economic growth however point at negative impacts of this on the future human beings life
- Health approach, where health can be defined as the state of complete physical, psychological and social well-being, not only as the absence of any disease or defect. Health is influenced by internal (impossible to influence) and external (possible to influence) influences
- Environmental approach where in most cases this approach is observed in relation with the health status of the population. Here we observe the quality of the air, potable water, noise level or the effect of foreign matter substances on the human body, e.g. from food

chains and similar then it is possible to define the sets of input variables (indicators) of the mentioned approaches that influence the QL.

In our case they are selected "objective" indicators of health and environmental approach, that are available in publicly available and accessible information resources, e.g. Czech statistical office database and similar.

When processing the QL indicators a number of mathematical methods can be used from one dimensional to multidimensional statistical methods, artificial and computational intelligence methods, decision trees, rough sets and the method of case based reasoning.

Real data from 1998 to 2007 coming from the Czech Statistical Office [18], [19], [20], the Institute of Health Information and Statistics of the CR, the Regional Information Service, the Czech Hydrometeorological Institute - Air Quality Control Division [21] were used for creation of QHS model.

Input matrix contains demographic attributes oriented to health [20], appendix attributes (number of physicians and hospitals) and selected environmental attributes. Used inputs have a wider significance and mutual relations, which are described below. The chosen attributes (parameters, indicators) are also selected in accordance with the probability of causes of death. For the first group of attributes (demographic and appendix) include following: attributes were analyzed cause of deaths based on International Statistical Classification of Diseases and Related Health Problem - diseases of the respiratory system, neoplasms, diseases of the circulatory system, diseases of the digestive system, external causes of morbidity and mortality [in numbers]; total deaths [in number]; number of physicians and hospitals; females and males life expectancy at birth [in years]; live and still births [in numbers].

Many factors (heredity, tobacco use, socio-economical status etc.) affect respiratory system diseases [22]. Malignant neoplasms (neolpasms) are also very numerous causes of death. The tumor may arise from various causes, mostly due to heredity, lifestyle, tobacco use, hormonal imbalances, environmental pollution or ionizing radiation [23]. The most common cause of death is considered cardiovascular system disease [24]. Cardiovascular system diseases are considered leading killer for both men and women among all racial and ethnic groups. There are many factors affecting cardiovascular diseases, most mentioned are, lifestyle behaviors or daily stress. Diseases of the digestive system are all of the diseases that affect the human digestive tract (pancreas, esophagus, stomach, small intestine, large intestine, liver). They are often caused by lifestyle, stress, poor diet, tobacco use. Numbers of physicians (doctors), establishments (hospitals) etc. characterize health care quality. Variations in life expectancy are mostly caused by differences in medical care, QL and diet. There are also differences in life expectancy between women and men, so life expectancy is usually calculated separately for males and females [25], [26]. The life expectancy at birth is highly sensitive to the infant mortality. The infant mortality is used as the indicator of a country's level development (poor countries have usually high rate of infant mortality).

The second group relates to: total environmental protection investment [in 10^3 CZK current prices] and main pollutant emissions - suspended particles, sulfur dioxide, nitrogen oxides, and carbon monoxide [in tons per year]. A description of attributes is in the Table 1.

Table 1 Definition of attributes

Attributes	Description of attributes
x_1	Cause of death - Diseases of the respiratory system
x_2	Cause of death - Neoplasms
x_3	Cause of death - Diseases of the circulatory (cardiovascular) system
x_4	Cause of death - Diseases of the digestive system
x_5	Cause of death - external causes of morbidity and mortality (suicides etc.)
x_6	Total deaths
x_7	Number of physicians
x_8	Number of hospitals
x_9	Females life expectancy at birth
x_{10}	Males life expectancy at birth
x_{11}	Live births
x_{12}	Still births
y_1	Total environmental protection investment
y_2	Suspended particles (PM_{10})
y_3	Sulfur dioxide (SO_2)
y_4	Nitrogen oxides (NO_x)
y_5	Carbon monoxide (CO)

We work with the model in two steps. In the first step a cluster analysis (CA) is used for linguistic level definition of QHS. CA [27], [28], [29], [30], [31], [32], [33], [34] is an exploratory data analysis tool for solving classification problems. The object is sorted into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. The task of clustering is then to divide the set of objects into the disjunctive clusters. The decision making about the object clustering in cluster is realized on the basis of the similarity by application of metric. The CA distinguishes hierarchical and non-hierarchical methods. The basic division of methods is mentioned for instance in [28], [30] and application in [35], [36], [37], [38]. Selected health (demographic) attributes are inputs for CA.

In the second step we apply selected algorithms of decision trees (DTs) for classification model creation. DT [28], [30], [34] is a predictive hierarchical model of decisions and their consequences, which can be used to represent both classifiers and regression models. DTs are frequently used in fields such as finance, risk management, marketing, medicine, etc. More examples are in [27], [31], [39]. Chosen environmental attributes and CA output (clusters) are inputs for DTs.

III. Problem Solution

Basic scheme of QHS modeling is in Fig. 1. The real data matrix **M**(140×17) represents model input. Partial outputs are presented by three linguistic levels of QHS – low, middle and high on the basis of $x_1, x_2, \ldots, x_{12}$. Final output is classification of QHS level by the decision trees on the basis of independence environmental attributes $y_1, y_2, \ldots, y_5$ (see Table 1).

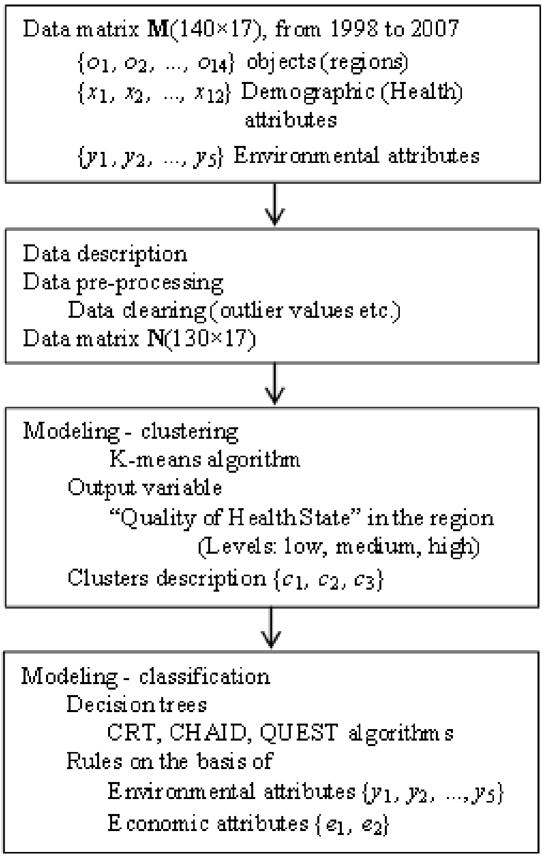


Fig. 1 Modeling of Quality of Health State in regions

In data pre-processing phase we eliminated 10 objects with outlying values (it concerned values about the CR capital - Prague) and recalculated values of selected attributes per 10^4 inhabitants. An existence of n objects is an initial condition for the usage of CA. The region in the years is object of the clustering (13 CR regions per 10 years). Every r-th object $\boldsymbol{o}_r$ is described by p and s characteristics (attributes, indicators). We can describe every object by the following vector:

$$\boldsymbol{o}_r = (x_{r1}, x_{r2}, \ldots, x_{rp}, y_{r1}, y_{r2}, \ldots, y_{rs}) \qquad (2)$$

where $r = 1, 2, \ldots, 130$; $p = 1, 2, \ldots, 12$; and $s = 1, 2, \ldots, 5$. The input set of the objects is possible to write in a formula of objects matrix **N**(130×17).

In the first step of modeling it was necessary to create clusters "levels of QHS in the regions" and to use them as output variable in the second step of modeling (by creation of classification model). For creation of clusters we used following attributes: $x_1, x_2, \ldots, x_8, x_{11}$ and x_{12} where they are in number per 10^4 inhabitants, and x_9 and x_{10} are in years. The K-means algorithm of CA is used. It is computed three clusters. Average values of attributes in clusters are in Table 2.

Table 2 Average values of attributes in clusters

Attributes	Clusters		
	c_1	c_2	c_3
x_1	0.026	0.058	0.045
x_2	0.142	0.321	0.252
x_3	0.282	0.654	0.478
x_4	0.022	0.054	0.040
x_5	0.034	0.075	0.060
x_6	0.539	1.247	0.929
x_7	0.186	0.417	0.271
x_8	0.001	0.002	0.002
x_9	78.638	78.555	76.900
x_{10}	72.236	71.766	70.164
x_{11}	0.495	1.093	0.851
x_{12}	0.001	0.003	0.003

In the comparison process of attribute values (see Table 2), it means their positive and negative feature, we can define the output variable QHS (for example "positive" are high values of attributes x_9, x_{10} and x_{11} , and "negative" are high values of attributes x_1, x_2, x_3, x_4 and x_5). The created clusters will be given information about "health state" – an indicator in a given territory.

Partial outputs variable - QHS for cluster c_1 (90 objects) is "High", for cluster c_2 (29 objects) is "Low" and for cluster c_3 (11 objects) is "Middle". It depends on positive and negative feature of attributes. For example Fig. 2 and Fig. 3 show graphical representations of attributes y_2 (suspended particles PM_{10}) and y_5 (carbon monoxide) in tons per year in the CR regions by designed QHS levels (low, medium, high). There is not extreme value for Moravskoslezsky region (y_5 is 145 666.18) in Fig. 3. After creation of clusters we focused on modeling by environmental and economical data.

A. Modeling by environmental data

In this step of modeling we dealt with classification of QHS for tree linguistic levels. We focused on algorithms based on DTs in the process of creating classification models for the solution of the mentioned task. We used the following algorithms (methods): Classification and Regression Trees (CRT), Chi-square Automatic Interaction Detection (CHAID), and Quick Unbiased Efficient Statistical Tree (QUEST). They belong to the most widely known DT algorithms [35]. A more detailed description of the algorithms is in [30].

By models creation we used these input environmental attributes $y_1, y_2, \ldots, y_5$ (see Table 1); and output attribute is represented by "levels of QHS in the regions". We composed set of objects on a training set (67% of objects) and a testing set (33% of object). Accuracy for training set A_{TRAIN} , accuracy for testing set A_{TEST} and number of created rules N_R is in

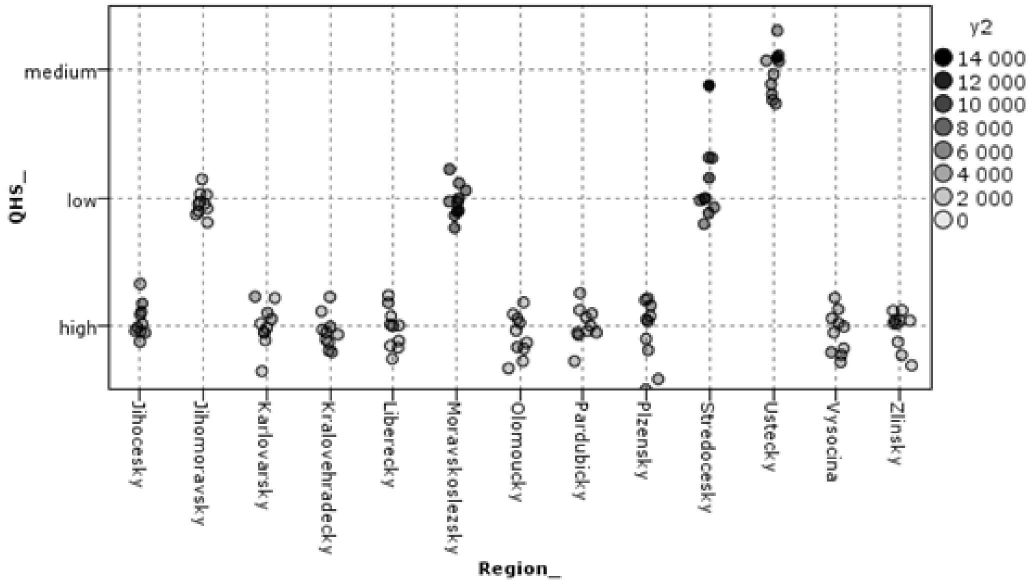


Fig. 2 Representation of attribute y_2 (y2) in regions (Region_) by designed QHS levels (QHS_)

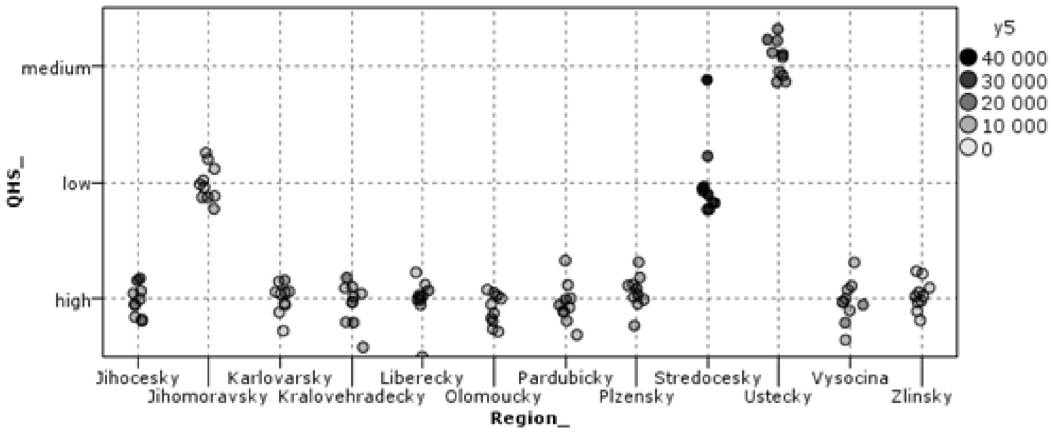


Fig. 3 Representation of attribute y_5 (y5) in regions (Region_) by designed QHS levels (QHS_)

Table 3.

We can see that the difference between results of training and testing set are considerably (for CRT it is 14.63%; for CHAID it is 7.32% and for QUEST it is 8.64%; it means that from view of models accuracy the CHAID and QUEST algorithms give similar results but differ in number of rules.).

Table 3 Accuracy of classification and number of classification rules

Algorithms	A_{TRAIN}	A_{TEST}	N_R
CRT	100%	85.37%	5
CHAID	100%	92.68%	8
QUEST	98.88%	90.24%	5

Example of five rules achieved by QUEST algorithm is the following:

- IF y_4 <= 27 711.53 AND y_3 <= 57 760.66 AND y_1 <= 1 518 954.13 AND y_5 <= 68 912.75
  THEN QHS level of region is "high".
- IF y_4 <= 27 711.53 AND y_3 <= 57 760.66 AND y_1 <= 1 518 954.13 AND y_5 > 68 912.745
  THEN QHS level of region is "low".
- IF y_4 <= 27 711.53 AND y_3 <= 57 760.66 AND y_1 > 1 518 954.13
  THEN QHS level of region is "low".
- IF y_4 <= 27 711.53 AND y_3 > 57 760.66
  THEN QHS level of region is "medium".
- IF y_4 > 27 711.53
  THEN QHS level of region is "medium".

B. Modeling by environmental and economic data

In this step of modeling we used two new economic attributes e_1 (GDP in region per capita in 10^4 CZK) and e_2 (a number of registered unemployed persons in %). We dealt with question if GDP and unemployment have any impact on levels of QHS in the regions. Representations of attributes e_1 and e_2 in CR regions by levels of QHS are in Fig. 4.

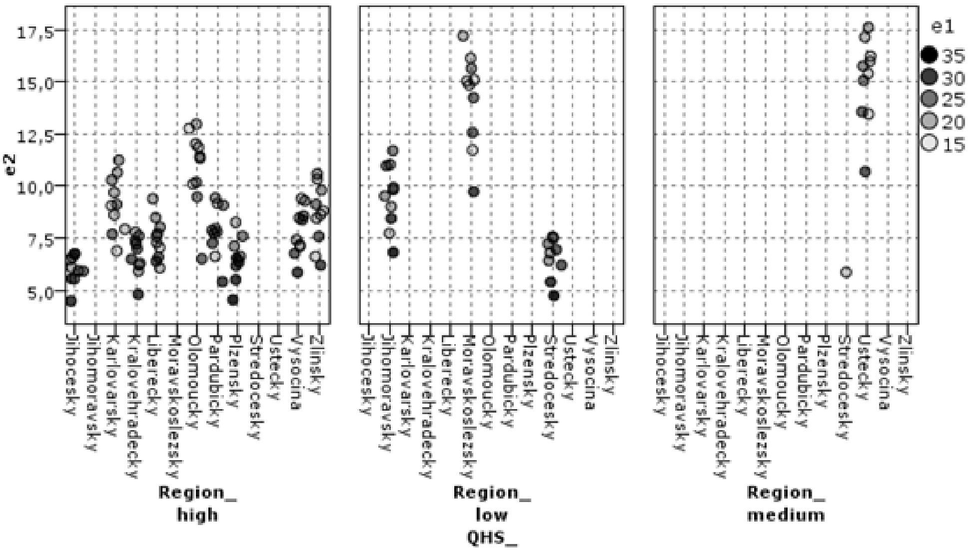


Fig. 4 Representations of attributes e_1 (e1) and e_2 (e2) in regions (Region_) for QHS level (high QHS_, medium QHS_ and low QHS_)

In Fig. 4 we can see that for example "Ustecky" region belongs to regions with high unemployment (e_2) and rather low values of GDP (e_1) and QHS in this region is "medium". Regions with "high" level of QHS belong to regions with rather higher values of GDP (e_1).

However, by using of above mentioned environmental y_1, y_2, …, y_5 and economic attributes e_1, e_2 similar results of modeling it was achieved (algorithm CRT: A_{TRAIN} is 100%, A_{TEST} is 92.68%; algorithm CHAID: A_{TRAIN} is 100%, A_{TEST} is 92.68%; algorithm QUEST: A_{TRAIN} is 98.88%, A_{TEST} is 90.24%; it means that better result of classification by use of economic attributes (difference is 7.31%) is only achieved by CRT algorithm).

Only this algorithm CRT used economic attributes e_1 and e_2 by creation of rules. Rules are the following:

- IF y_1 <= 1 798 531 AND e_2 <= 13.9 THEN QHS level of region is "high".
- IF y_1 <= 1 798 531 AND e_2 > 13.9 AND e_1 <= 211 470 THEN QHS level of region is "low".
- IF y_1 <= 1 798 531 AND e_2 > 13.9 AND e_1 > 211 470 THEN QHS level of region is "medium".
- IF y_1 > 1 798 531 AND y_3 <= 51 424.9 THEN QHS level of region is "low".
- IF y_1 > 1 798 531 AND y_3 > 51 424.9 THEN QHS level of region is "medium".

The designed classification models were created in Clementine ver. 10.0. [40], [41].

IV. Conclusion

Modeling of QHS level in regions offers various ways of processing. We dealt with application of CA and selected algorithms of DTs. By K-means algorithm we can sort regions in CR on the basis of health (demographic) attributes to these three clusters: High level of QHS (cluster c_1) is in Plzeňský, Královéhradecký, Pardubický, Zlínský, Jihočeský, Karlovarský, Liberecký, Vysočina, and Olomoucký region; Middle level of QHS (cluster c_3) is in Ústecký region; and Low level of QHS (cluster c_2) is in Středočeský, Jihomoravský, and Moravskoslezský region.

For classification of QHS level on the basis of environmental attributes we used CRT, CHAID and QUEST algorithms. By accuracy rate the best result of classification was achieved by CHAID algorithm (92.68%). On the basis of classification results we focused on the data collection and we searched other new possibilities how to specify QHS level in regions. We created new models based on economic and environmental data. By use of economic and environmental attributes results of CRT algorithm were only improved (difference between model created by environmental attributes and model created by economic and environmental attributes was 7.31%).

In future, it would be appropriate to become occupied with the analysis of the environment pollution resources in monitored localities.

References

[1] B. S. Blanchard, and W. J. Fabrycky, *System Engineering and Analysis*. 3rd Edition, New Jersey: Prentice Hall, 1998 (1990, 1981).

[2] B. S. Blanchard, *System Engineering Management*. 3rd Edition, New Jersey: John Wiley and Sons, Inc., 2004.

[3] E. Turban, J. A. Aronson, and T. P. Liang., *Decision Support Systems and Inteligent Systems*. Upper Saddle River: Prentice Hall, 2005.

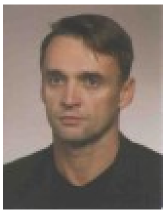
[4] P. Jirava, J. Mandys, M. Kasparova, and J. Krupka, "System Approach to Determinants of Quality of Life within a Region," *WSEAS Transaction on Systems*, vol. 9, pp. 243-252, 2010.

[5] J. Křupka, M. Kašparová, and P. Jirava, "Modelování kvality života pomocí rozhodovacích stromů," *Economics and Management*, vol. 13, no. 3, pp. 130-146, 2010. (in Czech)

[6] P. Mederly, J. Topercer, and P. Nováček, *Indikátory kvality života a udržitelného rozvoje : kvantitativní, vícerozměrný a variantní přístup*. Prague: Charles University UK FSV CESES, 2004. (in Czech)

[7] M. Hersh, *Mathematical modelling for sustainable development*. Berlin: Springer Verlag, 2005.

[8] J. Křupka, J. Svobodová, M. Augustinová, and J. Mandys, "Systémový přístup k problematice modelování kvality života," in *Proc. Conf. Aktuální otázky sociální politiky – teorie a praxe 2010*, Pardubice, CR: University Press, 2010, pp. 52-57. (in Czech)
[9] D. Akranavičiūtė, and J. Ruževičius, "Quality of Life and its Components' Measurement," *Engineering Economics* [Online]. 2007, no. 2, Available: http://web.ebscohost.com/ehost/pdf?vid=4&hid=2&sid=199fd247-f38c-4e61-9936-3aff1c300690%40sessionmgr3
[10] D. Phillips, *Quality of Life: Concept, Policy and Practice*. London: Routledge, 2006.
[11] M. Rapley, *Quality of Life Research: A Critical Introduction*. London: SAGE, 2003.
[12] Křivohlavého, J. *Psychologie zdraví*. Praha: Portál, 2009. (in Czech)
[13] Cakirpaloglu, P., *Psychologie hodnot*. Olomouc: Votobia s.r.o., 2004. (in Czech)
[14] Šebek, M., "K vymezení pojmu hodnota v psychologii," *Československá psychologie*, vol. 17, no. 3, pp. 252-258, 1973. (in Czech)
[15] Kavalíř, P., "K problematice výzkumu hodnot a hodnotových preference," *Československá psychologie*, vol. 49, no. 4, pp. 333-340, 2005. (in Czech)
[16] Hnilica, K., "Vlivy socializace a osobnosti na genezi konzervativních a liberálních hodnot," in *Kvalita života a zdraví*, J. Payne, Ed. Prague: Triton, 2005. (in Czech)
[17] Vašina, B., *Psychologie zdraví*. Ostrava: Ostravská univerzita v Ostravě, 1999. (in Czech)
[18] Czech Statistical Office, *Demographic Yearbook of the Districts of the Czech Republic 1998 – 2007, Code: 4034-08* [Online]. 2008. Available: http://www.czso.cz/csu/2008edicniplan.nsf/engpubl/4034-08-1998___2007
[19] Czech Statistical Office, *Statistical Yearbook of the Czech Republic 2008, Code: 0001-08* [Online]. 2008. Available: http://www.czso.cz/csu/2008edicniplan.nsf/engpubl/4034-08-1998___2007
[20] Czech Statistical Office, *Demographic Yearbook of the Regions of the Czech Republic 1998 - 2007, Code: 4027-08* [Online]. 2008. Available: http://www.czso.cz/csu/2008edicniplan.nsf/engpubl/4027-08-1998___2007
[21] Czech Hydrometeorological Institute, *Environmental Protection Expenditure in the Czech Republic 2008* [Online], 2008. Available: http://www.czso.cz/csu/2007edicniplan.nsf/p/2005-07
[22] M. Weires, J. L. Bermejo, K. Sundquist, and K. Hemminki, "Socio-economic status and overall and cause-specific mortality in Sweden," *BMC Public Health*, 2008, vol. 8, 340. [Online]. Available: http://www.biomedcentral.com/content/pdf/1471-2458-8-340.pdf
[23] *National Cancer Institute*, 2010. Available: http://www.cancer.gov/cancertopics/wyntk/cancer/page4
[24] World Health Organization, *Annex Table 2: Deaths by cause, sex and mortality stratum in WHO regions, estimates for 2002*. 2004. Available: http://www.who.int/entity/whr/2004/annex/topic/en/annex_2_en.pdf. Retrieved 2010-10-7
[25] J. C. Riley, *Rising life expectancy: a global history*. Cambrige, UK: Cambridge University Press, 2001.
[26] J. C. Riley, *Poverty and life expectancy: the Jamaica paradox*. Cambrige, UK: Cambridge University Press, 2005.
[27] P. Guidici, *Applied Data Mining: Statistical Methods for Business and Industry*. West Sussex: Wiley, 2003.
[28] J. Han, and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers, 2001.
[29] A. Lukasová, and J. Šarmanová, *Metody shlukové analýzy*, Prague: SNTL Nakladatelství technické literatury v Praze, 1985. (in Czech)
[30] O. Maimond, L. Rokach, *The Data Mining and Knowledge Discovery Handbook*, New York: Springer, 2005.
[31] D. Pyle, *Business Modeling and Data Mining*, San Francisco, KA: Morgan Kaufmann Publishers, 2003.
[32] H. Fu, "Cluster Analysis and Association Analysis for the Same Data," in *Proc. 7th WSEAS Int. Conf. Artificial Intelligence, Knowledge Engineering and Data Bases*, Cambridge, 2008, pp. 576-581.
[33] D. A. Freedman, *Statistical Models: Theory and Practice*. New York, NY: Cambridge University Press, 2005.
[34] *Electronic Statistics Textbook*, StatSoft, Inc., Tulsa, OK, 2010. [Online]. Available: http://www.statsoft.com/textbook.
[35] I. H. Witten, and E. Frank, *Data Mining: Practical Machina Learning Tools and Techniques*, 2nd Editions, San Francisco: Morgan Kaufmann Publishers, 2005.
[36] M. Kašparová, and J. Křupka, "Classification and Prediction Models for Internal Population Migration in Distrists," *WSEAS Transaction on Systems*, vol. 5, pp. 1540-1547, 2006.
[37] J. Křupka, P. Jirava, J. Mandys, and M. Kašparová, "Quality of Life Investigation Case Study in the Czech Republic," in *Proc. 6th Int. Conf. on Fuzzy Systems and Knowledge Discovery*, Tijanin, China: IEEE Computer Society, vol. 1, 2009, pp. 264-268.
[38] M. Kasparova, J. Krupka, and J. Chylkova, "Heavy Metals Contamination Analysis in Selected Czech Localities by Cluster Analysis," *Recent Advances in Environment, Ecosystems and Development*, Athens New York: WSEAS Press, 2009, pp. 240-581.
[39] J. Křupka, J. Mandys, and P. Jirava, "Approaches for the Comparison of the Quality of Life Investigation," in *5th EA-SDI Conf. Proc. Environmental Accounting – Sustainable Development Indicators*, Prague, CR, 2009.
[40] *Clementine® 12.0 User's Guide*, SPSS, Inc., Chicago, IL, 2008.
[41] *Clementine® 7.0 User's Guide*, SPSS, Inc., Chicago, IL, 2002.



Jiří Křupka was born in Prostějov (CR) in 1962. He graduated from the Military Technical University in Liptovský Mikuláš (Slovakia) in 1985. From 1985 till 1990 he worked in the Department of Technical Support System's and Automation in the Air Defense. From 1990 till 2004 he worked as a lecturer, a senior lecturer, and vice-dean for education at the Faculty of Air Defense at the Military Academy in Liptovský Mikuláš. There he finished his doctoral thesis in 1995 and habilitated in 1997. Since 2004 he is working as associated professor and head of Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice (CR).

Assoc. Prof. Křupka has published parts of book and a number of papers concerning with fuzzy decision, fuzzy control, case based reasoning, and rough set theory. Nowadays he is focusing on modeling of environmental and social systems.

Miloslava Kašparová was born in Klatovy (CR) in 1976. She is a senior lecturer at Institute of System Engineering and Informatics at Faculty of Economics and Administration at the University of Pardubice. There she received the Master's degree in economy in 2000, and in 2005 she finished the doctoral thesis in the field of informatics in public administration. She has dealt with the modeling of processes in the public administration and nowadays she focuses on an application of selected data mining methods in environmental and social systems.

Dr. Kašparová has published papers concerning an application of data mining methods in various areas.

Jan Mandys was born in 1979. He graduates of Social work with the advisory focus of University of Ostrava, Faculty of Arts, Department of Psychology and Social Service (2003). From the October 2006 he studies Doctoral study programme Social Counselling and Social Service Management of University of Ostrava (CR).

He works as Assistant Professor at University of Pardubice, Faculty of Economics and Public Administration, Institute of Public Administration and Law (since October 2003). Before that, he worked as Social curator and counsellor for Romany people of Municipality of Pardubice, Department of social affairs (January 2007 – August 2008), Social worker and manager in the non-profit organization o. s. SKP-CENTRUM (October 2003 – December 2006) and during study he worked two years as Social worker and manager in the non-profit organization o. s. Nový Prostor. His publications are mainly related to homelessness, quality of life and social policy of public administration. Publications: Methodology of Working with people released from prison (Prague, CR: Evropské sociálně zdravotní centrum Praha, o.p.s., 2008);

Implementing various methodological approaches to research of quality of life as possible support methods of public administration decision making (Pardubice, CR, University of Pardubice, 2009); Determinants of Quality of Life within a Region (WSEAS PRESS, 2009). He professionally focuses on social pathology, sociology, social psychology and social politics

Mr. Mandys is a member of the working group of community planning of social services of Municipality of Pardubice and medium-term plan of social services of Regional Authority of the Pardubice Region. In the past he was a member of the national organizations of non-governmental organizations and a member of the Health and Social Commission of Union of Towns and Municipalities of CR.

Pavel Jirava was born in Hradec Kralove (CR) in 1975. Has received the Master's degree in economy, University of Pardubice, 1998, and PhD. degree in System Engineering and Informatics, University of Pardubice, 2007.

From 2004 he is senior lecturer at the University of Pardubice, Faculty of Economics and Administration. His principal research interests are soft computing, data mining, rough sets, their application in the area of sustainable development and related fields. He has published more than 20 conference papers, 8 papers in journals, 5 book chapters. Three selected publications: Classification model based on rough and fuzzy sets theory. Author(s): Jirava, P., Křupka, J. Puerto de la Cruz, Spain: CIMMACS '07: Proceedings of the 6th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics – WSEAS Press. Published: 2007; Case-Based Reasoning Model in Process of Emergency Management. Author(s): Křupka, J., Kašparova, M., Jirava, P. Kocierz, Poland: Man-Machine Interactions – Springer Verlag. Published: 2009; Air Quality Modelling by Means of Rough and Fuzzy Sets: Modelling of Selected areas of Sustainable Development by Artificial Intelligence and Soft Computing - Regional level. Jirava, P., Křupka, J., Kašparová M. Prague, Czech Republic: GRADA Publishing a.s. Published: 2009.

Dr. Jirava was also principal investigator of 2 research projects and co-investigator of 3 research projects.