Igor V. Hut

Researcher University of Belgrade Faculty of Mechanical Engineering NanoLab, Biomedical Engineering

Jelena S. Muncan

Assistant University of Belgrade Faculty of Mechanical Engineering NanoLab, Biomedical Engineering

Branislava D. Jeftic

Researcher University of Belgrade Faculty of Mechanical Engineering NanoLab, Biomedical Engineering

Sanja T. Dogramazi

Associate Professor Bristol University Department of Mehanical Engineering, London, UK

Lidija R. Matija

Associate professorUniversity of Belgrade Faculty of Mechanical Engineering NanoLab, Biomedical Engineering

1. INTRODUCTION

Identifying a quality of water used for human consumption has always been one of the most serious challenges. Standards have been developed by national and international organizations to define a quality of water. Most of these standards set upper limits for physical parameters, chemical constituents and presence of microorganisms.

Mineral water is consumed as a drink beneficial for human health and is used for therapeutic purpose. Mineral water is often considered to be safer and healthier drink than tap water. However, the quality and composition of mineral water vary with its origin and require careful monitoring. The quality of drinking water, is usually considered in terms of the concentrations of different solutes, cations and anions, the ratio of certain ions etc., with regard to human organism functioning. Little is known about the structure of the water itself, even though with approximately 72%, of the water content in the human body, it is the most abundant chemical and plays a central role in the regulation of cell volume, nutrient transport, waste removal and thermal regulation [1].

This aspect of water needs to be addressed especially since the market today features water brands offering alleged beneficial effects and typically described as 'revitalized', 'magnetized', or exposed to other external factors which can influence structuring of liquid water. Opto-magnetic imaging spectroscopy

Received: March 2014, Accepted: May 2014 Correspondence to: Igor Hut Faculty of Mechanical Engineering, Kraljice Marije 16, 11120 Belgrade 35, Serbia E-mail: ihut@mas.bg.ac.rs doi:10.5937/fmet1403256H © Faculty of Mechanical Engineering, Belgrade. All rights reserved

Multivariate Analysis and Self Organizing Feature Maps Applied for Data Analysis of Opto-Magnetic Spectra of Water

To obtain new knowledge about structure of liquid water and interaction between constituents in water and water molecules a new approach has been attempted using Opto-magnetic imaging spectroscopy (OMIS), a method based on a light-matter interaction. Opto-magnetic imaging spectroscopy is a novel method which takes into account a ratio of electrical and magnetic forces of chemical bonds, and therefore collects data of both classical and quantum actions of water molecules and other constituents. Here, we used OMIS combined with techniques of multivariate analysis and neural networks to extract data from spectra of different waters. We have investigated this method for characterization and discrimination of different waters with special interest in para- diamagnetic properties which can give clues about organization of water molecules. It is shown that the use of OMIS together with multivariate techniques and neural networks approach can be proved as a valuable asset in characterizing water from the aspect of its structural organization.

Keywords: Self organizing feature maps, Multivariate data analysis, Water, Water Clustering, Opto-magnetic spectroscopy.

(OMIS) is a novel tool successfully applied in characterization of different materials [2, 3, 4, 5], and can indirectly provide information about organization of water molecules and water clusters by measuring paramagnetic and diamagnetic properties of water. [6]. Therefore, OMIS can be proved to be a valuable tool in assessing water quality for different health benefits.

In this study Opto-magnetic imaging spectroscopy was used to enquire about differences between several types of water: tap water, commercial mineral waters – carbonated and non-carbonated and water solutions. Water available on Serbian market – Knjaz Milos, Aqua Viva [3], Zlatibor – carbonated and not carbonated [4] has been investigated. These are popular water brands reported to have beneficial health effects.

Since OMIS gives results in a form of spectra, data mining methods are necessary to extract information from the spectra, and this research is based on using multivariate analysis and neural networks. The present work describes the results achieved and the obstacles encountered on the way.

2. OPTO – MAGNETIC IMAGING SPECTROSCOPY IN CHARACTERIZATION OF WATER

2.1 Material

Physico-chemical properties of the analysed waters and water solutions are presented in Table1. along with some typical concentrations of the water ions.

Commercial mineral waters were obtained at a local supermarket; etanolum dilutum and aqua purificata were obtained at a local pharmacy. The samples of tap water were stored in a polyethylene bottle rinsed thoroughly with pure water prior to collecting. All mineral water brands were in original plastic bottles with the capacity of 0.51 closed with plastic screw caps. Manufacturers' labels on the bottles were used as a source of basic information about particular water types, as well as the data from the manufacturers' websites [7, 8, 9]. Physico-chemical properties of tap water were obtained from the last published data on European water [10], and from public health institutions in Belgrade [11]. Aqua Purificata Sterilisata, 500ml (Pharma product) and Etanolum dilutum, 70%, 500ml (Hemofarm) were obtained at the local pharmacy.

2.2 Method

Opto-magnetic imaging spectroscopy is a method based on light-matter interaction (For more detailed description of the method see references [2-6]).

The device used for taking digital images is OMIS WP-B53. It comprises a digital camera (Canon XS 105, Cannon Inc.) in a specially designed housing with an additionally placed two LED systems (six light emitting diodes per system) arranged in a circle in front of an objective for illuminating surface of a sample. First LED system illuminates sample by providing incident white diffuse light under the angle of 90°. The second LED system provides incident white diffuse light under the angle of 53° - the angle under which light reflected from the water surface will be completely polarized (Brewster angle for water) [3].



Figure 1. Selected pairs (white and polarized) of digital images of water (1) Knjaz Milos, (2) Zlatibor non-carbonated, (3) Belgrade tap water (right).

Since reflected polarized light contains only electrical component of light-matter interaction, the difference between reflected white light (electromagnetic nature) and reflected polarized white light can serve as a measure of magnetic properties of matter.

For each water sample, at least 10 pairs of white and polarized digital images were taken. The images used are observed in RGB colour space, but only red and blue colour data channels were used for both white diffuse light (W) and reflected polarized white light (P). An algorithm for data analysis is based on chromaticity diagram called Maxwell's triangle and spectral convolution operation according to ratio (R-B) & (W-P). The abbreviated designation means that Red minus Blue wavelength of white light and reflected polarized light are used in spectral convolution algorithm to calculate data for opto-magnetic fingerprint of matter.

Using this algorithm for spectral convolution, 10 pairs of white and polarized images of water sample, give 10 spectra per water sample. In total 71 spectra were acquired.

2.3 Multivariate data analysis

All multivariate spectral analysis was carried out by Pirouette ver. 4.0 (Infometrics, USA) software program.

Principal component analysis (PCA) is a wellknown statistical method for reducing the dimensionality of data sets [12,13]. Its operation can be thought of as revealing the internal structure of the data in a way which best explains the variance in the data. The new dimensions, principal components – PCs – are built taking into account the maximum variance of data and the requirements about an orthogonal space. The number of PCs is much lower than the number of original variables, mainly in spectral analysis, due to the linear combination of the original variables in order to form the PCs thus removing co-linearity between variables. The results of a PCA are usually discussed in terms of component scores and loadings.

Soft independent modelling of class analogy (SIMCA) employs PCA for the construction of mathematical models for each class to be analysed [14]. It is a supervised pattern recognition technique considered the key chemo-metric approach for classification.

This technique enables classification of samples into an already existing group, assigning new objects to the class to which they show the largest similarity. SIMCA is strongly based on PCA, because each class is defined by an independent PCA, taking into account the optimal number of PCs for each class, which is endowed with a specific data structure.

2.4 Self-organizing feature maps

Self-organizing feature map (SOFM) is a type of artificial neural network that uses unsupervised learning as training method to produce two-dimensional data presentation from high-dimensional input data so that data structure can be validated visually [16]. The SOFM can be used not only to cluster input data, but also to explore the relationship between different attributes of input data. It could be said that SOFM consists of two layers: an input layer formed by a set of n-dimensional input vectors, and an output layer formed by units (neurons) arranged, usually, in a two-dimensional grid or lattice. Each neuron practically represents an index of n-dimensional weight vector which corresponds to observed input vector. The most common definitions of the neuron neighbourhood are: linear, square and hexagonal. Network parameter that links the input to the output neurons is the weight. In SOFMs, winner neuron and neurons in the neighbourhood of the winner have similar weights [16].

There are two modes of training in SOFM networks: recursive and batch mode. The difference between these two modes is weight adjustment. In recursive mode, weight adjustments are made after each input pattern is presented, and in batch mode one weight adjustment is made after an epoch. Self-organizing feature maps allow nonlinear projection of the data. Training is usually performed in two phases: ordering and convergence (tuning). In the ordering phase, learning rate and neighbourhood size are reduced with iterations until the winner or a few neighbours around the winner remain. In this phase, a topological ordering of the weight vectors takes place. In the ordering phase, the map is trained and in the convergence phase it is fine-tuned [16].

In SOFMs, the distance between neuron weights is used as a visualization tool because it can highlight different cluster regions in the map. The average of the distance to the nearest neighbours is called unified distance, and the matrix of these values for all neurons is called the U-matrix. In U-matrix, lighter regions represent smaller distances between the neighbouring vectors and darker regions represent larger distances between them. This kind of visualization can only be used to obtain qualitative information.

3. RESULTS AND DISCUSSION

All water samples show the most significant differences in the part of the spectra from 110nm to 140nm on wavelength difference axis. In Fig. 3 averaged spectra for all examined samples are presented. From this part of the spectra, the main spectral features – first positive peak and corresponding wavelength difference, first negative peak and corresponding wavelength, and second positive and negative peaks and corresponding wavelength differences – are extracted and used in further analysis (Example of extracted spectral features is shown on the averaged spectrum of Aqua Purificata water shown in Fig. 3).



Figure 2. Averaged spectra for all water samples.

Aqua Purificata - Average spectrum in magnetic domain [R-B]&[W-P]



Figure 3. Spectral features used in multivariate analysis and SOFM.

3.1 Self-organizing feature maps

Self-Organizing Feature Map (SOFM) used for spectral data analysis was constructed in the Matlab R 2010a (MathWorks, USA) programming environment. Input vectors were formed using the following data: first positive peak, first negative peak, second positive peak, and second negative peak for each sample. These vectors were arranged in columns of the input matrix. This way input matrix of dimensions 4×71 was formed. We used a hexagonal topology of 5×8 (40) neurons for the neuron layer of our SOFM. Training of the SOFM was performed in the batch mode, which is the default learning mode in the Matlab. Parameters used for the network training were:

- total number of steps (epochs): 2000;
- number of steps in the ordering phase: 100;
- initial neighborhood size: 5.

These parameters provided optimal behavior in terms of distribution of network through the input space (Fig.4).



Figure 4. SOFM weight positions (weight 1 – weight 2 plane). Green dots represent data points, blue dots represent weight vectors and lines represent neuron connectivity.

Results obtained by the use of unified distribution matrix (U-matrix) suggest clustering of data into two groups, represented by the lighter region on the right of U-matrix and the darker region on the left of U-matrix (Fig.5). Lighter region of U-matrix indicates that the corresponding weights are closer together in this region, while darker colors on the left side of neighbor distance figure indicate that data points in this region are farther apart [17].



Figure 5. Unified distribution matrix (U-matrix), for used input data set.

This grouping of data points and accordingly network weights can also be observed in the *weight1weight2* plane of weight space (Fig. 4). Inspection of individual weight planes suggests that inputs 1 and 3 corresponding to first positive peak and second positive peak might be correlated to a higher degree (Fig.6).



Figure 6. Individual weight planes.

3.2 Principal component analysis

PCA technique was used to investigate the overall variation of data and reduce the number of dimensions present in the data matrix. PCA modelling was developed using a cross-validation (leave-10-out), and 3 outliers were removed according to the Mahalanobis distance [14].

The data dimension was reduced on three principal components (PC1, PC2 and PC3) explaining more than 99 % of variance (Table 2).

 Table 2. Variances in the spectral features data explained by the first three vectors

	Variance	Percent	Cumulative
Factor1	7243692.50	91.19	91.19
Factor2	542353.69	6.83	98.02
Factor3	93451.47	1.18	99.20

FME Transactions

The pattern of each PC presented as its loading in Fig.6 for the data set revealed the variables where main spectral variations occurred. These plots show the important variables in the respective pattern of each PC. From Fig.7 it is evident that the major variations in data occur in PC1 direction. It is also worth mentioning that the scatter plot revealed the linear correlation between first two PCs. Since all data are located on the positive side of the PC1 vector it is clear that the major variations are correspondent to variables on the positive side of the Loading plot (Fig.7b right half-plane). However it is more interesting to analyse second PC vector, which divides waters spectral data in positive and negative part of the scatter plot. It is evident that the negative part of the PC2 corresponds only to variables Peak2- and Peak1+, i.e. the intensities of the second largest negative peak and the first largest positive peak. Aqua Viva, Aua Purificata and Etanolum dilutum (70% water) spectral data oscillate around zero value and show some normal deviations. This is normal phenomenon since the number of hydrogen bonds formed by water molecules in liquid water fluctuates with time and water molecular species are constantly forming and breaking. This dynamic is highly dependent on temperature and pressure [15].



Figure 7a. Scatter plot.





Figure 7b. Loadings plot.

The spectral data for non-carbonated Zlatibor water are entirely located on the negative part of the PC2, while for Belgrade tap water, Knjaz Milos water and carbonated Zlatibor water are located in the positive part. The interesting conclusion can be drawn out that the water with presence of gasses : Belgrade (Chlorine, even though contrary to official information presented in Table1, presence of large amounts of chlorine is evident on the Fig.1), Knjaz Milos and carbonated Zlatibor water (carbon dioxide) which molecules are large and have large molecular weight are located in the positive part of the scatter plot while non-carbonated Zlatibor water (oxygen) with small molecules of free oxygen is entirely on the negative part. Since water molecules tend to organize around molecules of gasses in so called water shells, it is our suggestion that the second PC vector reveals the information about the size of water clusters present in the water. The distribution of Aqua Viva, Aqua Purificata and Etanolum dilutum data further supports this conclusion, since these are low mineral water, pure water and water solution and they are almost equally distributed in both parts of the spectral plot, thus suggesting that the PC1 vector explains normal variations in water dynamics: dynamics concerns only free water molecules or water clusters comprised of mainly water molecules, or water molecules around small solutes.

3.3 Soft modelling of class analogies

SIMCA modeling as a type of supervised classification was developed for each class (water type) using a crossvalidation method (leave 10 out). The results in Table 3. show the separation of classes (water types) according to Mahalanobis distance. The interclass distance (Mahalanobis distance) is in some cases less than 3 which illustrates relatively poor separation of classes. This indicated similarity of those water types (for example Knjaz Milos and carbonated Zlatibor water).

The Table 4. presents the results of prediction with created SIMCA model for each class (The numbers @No in Table4 is the number of PC factors used to model each class member).

In order to identify which variables – spectral features are most useful for discriminating between classes of water; the discrimination power is calculated and presented in Fig. 8 where it is shown that the variable which is most different in all water types is the intensity of the second largest positive peak on the wavelength difference axis of spectral plot.



Discriminating Power

Figure 8. Discriminating power of SIMCA model.

4. CONCLUSION

In this study neural networks non supervised classification technique together with multivariate

analysis was used to extract the useful information about water samples from the spectral data acquired using opto-magnetic spectroscopy. Since the method of Opto-magnetic imaging spectroscopy relies on finding the magnitude of paramagnetic and diamagnetic properties in relation to the size of water clusters [6], we expected to reveal some information about structural and organization characteristics of water molecules in analysed waters. This novel method for characterization of water could be proved to be very useful in covering these aspects of water molecules which contribute to water quality and are not in any way regulated at the present.

In the first step of data analysis we have used SOFM as an unsupervised method for discovering hidden clusters in data. This revealed existence of two clusters of data. One of the clusters representing data points in close proximity. Results obtained by this method also suggest relatively high correlation between first and second maximal positive peak. Usage of larger data set and addition of adequate components of input vectors, in this step of data analysis, would probably lead to more refined observations.

We further applied principal component analysis (PCA) and soft independent modelling of class analogy (SIMCA). The application of PCA revealed that the major changes in these waters spectra occur along direction of only 2 principal component vectors. We reach to a preliminary conclusion that the first PC vector which explains the most of the variation explains normal water dynamics – making and breaking of water cluster, while second PC vector explains the rest of variation which is related to water organization around molecules of gases. The changes along PC1 vector are probably due to exposure to environment and additional changes to water chemistries which can occur during the storage such as evaporation of dissolved gasses, precipitation of constituents or exposure to light.

The developed SIMCA discrimination models revealed that the analysed waters are different mostly in the intensity of the second positive peak.

This work opened a lot of questions that need to be answered. The use of OMIS together with multivariate techniques and neural networks approach can be proved to be valuable in characterizing water from the aspect of its structural organization. But to do this it is necessary to perform series of controlled experiments to identify exact correlation between the type of water molecular organization and demonstrated opto-magnetic properties of water. If this goal is achieved, the road for characterizing waters from the new – structural aspect with a simple and easily available method could be established.

ACKNOWLEDGMENT

This research has been partially funded by ministry of education and science of republic of Serbia, through projects III45009 and III41006.

Authors wish to express their sincere gratitude to Igor Koruga (senior researcher at DIA Systems, USA) for donation of the Pirouette ver. 4.0 (Infometrics, USA) software program.

REFERENCES

- [1] Schoeller, D.A.: Changes in total body water with age, American Journal of Nutrition Vol. 50: p. 1176-81, 1989.
- [2] Koruga, Dj., Tomic, A., Method and alorithm for analysis of light-matter interaction based on spectral convolution, US Pat. App. No. 61/061, 852, 2008, PCT/US2009/030347, Publ. No: WO/2009/089292, Publ. Date:2009-07-16.Stokes, J.: Production of Coated and Free-Standing Engineering Components using the HVOF (High Velocity Oxy-Fuel) Process, PhD thesis, School of Mechanical and Manufacturing Engineering, Dublin City University, Dublin, 2003.
- [3] Koruga, Dj., Miljkovic, S., Ribar, S., Matija, L., Kojic, D.: Water Hydrogen Bonds Study by Opto-Magnetic Fingerprint Technique, ACTA PHYSICA POLONICA A, Vol. 117, No. 5, pp. 777-781, 2010.
- [4] Muncan, J.: A comparative study of structure and properties of water by IR and Opto-magnetic spectroscopy, in: Book of Abstracts, The second scientific international conference on Water and Nanomedicine, Academy of Sciences and Arts of Republic of Srpska, 2011, Banja Luka, pp. 56-57.
- [5] Jeftic, B., Hut, I., Mladenovic, D., Muncan, J., Golubovic, Z., Sarac, D.: Characterization of solid, viscoelastic and liquid materials by Opto-magnetic spectroscopy, in Book of Abstracts, Thirteenth annual conference of the Materials Research Society of Serbia - YUCOMAT, 2011, Herceg Novi, Montenegro, pp. 136.
- [6] Koruga, Dj.:Paramagnetic/diamagnetic properties of water and aqueous solutions: classical/quantum approach for biomedical applications, in: Book of Abstracts of the International Conference on Water, Hydrogen Bonding Nanomaterials and Nanomedicine, 2011, Banja Luka, pp. 20-21.
- [7] Aqua Viva http://www.aquaviva.rs/aquaviva
- [8] Knjaz Milos a.d.; Available from: http://www.knjaz.co.rs
- [9] http://www.zlatibo2rvoda.rs/
- [10] Reimann C., Birke M.: Geochemistry of European Bottled Water, Gebrüder Borntraeger Verlagsbuchhandlung, Stutgart, Germany, 2010. ISBN 978-3-443-01067-6.
- [11] Environment quality of Belgrade in 2009 (in Serbian), R.M. M.Grubacevic, B.Glamocic, B.Bozovic, M.Tanaskovic, A.Popovic, Editor. 2010, City Institution for Public Health, Regional Environmental Center, City Administration Office: Belgrade.
- [12] Wold, S., Ebensen, K., and Geladi, P., Principal component analysis. Chemometrics and Intelligent Laboratory systems, Vol. 2 (1-3): pp. 37-52, 1987.

- [13] Brereton, R.G.: Chemometrics Data Analysis for the Laboratory and Chemical Plant., John Wiley & Sons Ltd, 2003.
- [14] Svante, W. and Michael, S.: SIMCA: a method for analyzing chemical data in terms of similarity and analogy in Chemometrics: Theory and Application, B. Kowalski, Editor, American Chemical Society. p. 243-282, 2003.
- [15] Roy, R., W.A.T., Bell, I., Hoover, M.R.: The Structure Of Liquid Water; Novel Insights From Materials Research; Potential Relevance To Homeopathy. Materials Research Innovations Online Vol. 9(4), 2005.
- [16] Kohonen, T.: Self-Organizing Map. Springer, Berlin, Heidelberg.1995. (Third, Extended Edition, 2001).
- [17] MathWorks, Matlab R2011b Documentation, Neural Network Toolbox, http://www.mathworks.com/ help/toolbox/nnet/ug/bss4b l-1.html.

ПРИМЕНА МУЛТИВАРИЈАНТНЕ АНАЛИЗЕ И САМО-ОРГАНИЗУЈУЋИХ МАПА ЗА АНАЛИЗУ ПОДАТКА ДОБИЈЕНИХ НА ОСНОВУ ОПТО-МАГНЕТНИХ СПЕКТАРА ВОДЕ

Игор В. Хут, Јелена С. Мунћан, Бранислава Д. Јефтић, Сања Т. Дограмази, Лидија Р. Матија

Како би се дошло до нових сазнања у вези структуре воде у течном стању, као и интеракција између растворених материја и молекула воде коришћена је нова метода базирана на употреби опто-магнетне имиџинг спектроскопије (ОМИС). У питању је метода базирана на интеракцији видљиве светлости и материје, која омогућава детекцију односа електричних и магнетних сила хемијских веза, те даје податке како о класичним тако и квантним дејствима између молекула воде и других материја у њој растворених. У овом раду су за екстракцију и анализу података, добијених употребом ОМИС, коришћене мултиваријантна анализа и неуронске мреже. Овим приступом смо покушали да извршимо карактеризацију и дискриминацију различитих вода са посебним интересовањем за њихове пара- и дијамагнетне особине које указују на самоорганизацију (кластеризацију) молекула воде. Показано је да употреба ОМИС паралелно ca техникама мултиваријантне анализе и неуронских мрежа може успешно бити употребљена за карактеризацију вода са аспекта њихове структуралне организације.

				Ca	Fe	К	Mg	Na	HCO ₃ -	Cl-	SO4 ²⁻	CO ₂	Dry content at 180°	Comment
No.	Sample type	рН	EC [µS/cm]	mg/l	mg/l									
1	Aqua Viva	7.62	504	90.1	< 0.05	2.0	14.3	10.5	320	12.0	26	-	370	*
2	Knjaz Milos	5.7	*	108.2	0.05	16.6	53.5	229.5	1281	10.7	13	min.3000	*	*
3	Zlatibor non carbonated	7.9	*	62.7	*	0.9	30.5	4.51	353	1.4	17.6	*	304	Naturaly dissolved O ₂ (9.8 mg/l)
4	Zlatibor carbonated	7.9	*	62.7	*	0.9	30.5	4.51	353	1.4	17.6	min.3000	304	Naturaly dissolved O_2 (9.8 mg/l)
5	Aqua Purificata	*												
6	Beograd	7.56	318	61.1	2.34	1.1	11.1	6.5	194	11.3	37.5		*	*
7	Etanolum dillutum	*												

Table 1. Investigated waters and concentrations of typical ions ("" no data available)

Table 3. Interclass (Machalanobis) distance

	CS1@5	CS2@3	CS3@4	CS4@3	CS5@5	CS6@5	CS7@2
1-Aqua Viva	0.00	3.27	10.66	2.68	19.38	268.05	1.48
2-Knjaz Milos		0.00	6.92	0.26	2.63	5.09	0.53
3- Zlatibor-non carbonated			0.00	7.44	10.40	53.71	3.97
4-Zlatibor-cabonated				0.00	2.27	5.81	1.03
5-Aqua Purificata					0.00	104.20	1.80
6-Belgrade tap water						0.00	7.04
7-Etanolum dillutum							0.00

Table 4. Results of prediction using SIMCA model

	Pred1@5	Pred2@3	Pred3@4	Pred4@3	Pred5@5	Pred6@5	Pred7@2	No match
Actual1	8	0	0	0	0	0	0	0
Actual2	0	10	0	2	0	0	0	0
Actual3	0	0	14	0	0	0	0	0
Actual4	0	6	0	5	1	0	0	0
Actual5	0	0	0	0	9	0	0	0
Actual6	0	0	0	0	0	8	0	0
Actual7	0	3	0	1	1	0	0	0