Chemometrics and Intelligent Laboratory Systems 118 (2012) 280–286

Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Surface water quality assessment using self-organizing maps and Hasse diagram technique

Tsvetomil Voyslavov *, Stefan Tsakovski, Vasil Simeonov

Group of Chemometrics and Environmetrics, Chair of Analytical Chemistry, Faculty of Chemistry and Pharmacy, University of Sofia "St. Kl. Okhridski", 1, J. Bourchier Blvd., 1164 Sofia, Bulgaria

ARTICLE INFO

Article history:
Received 14 March 2012
Received in revised form 22 May 2012
Accepted 23 May 2012
Available online 31 May 2012

Keywords:
Surface water
Quality assessment
Self-organizing maps
Hasse diagram technique

ABSTRACT

The present study deals with the important issue of assessing surface water quality by the use of advanced multivariate data treatment approaches like self-organizing maps of Kohonen (SOM) and Hasse diagram technique (HDT). The object of the study is the catchment of the transboundary Mesta River on Bulgarian territory. Long-term monitoring data (1990–2009) were collected from all sampling sites along the river flow involving all major surface water quality parameters. The clustering of the data by the use of SOM has helped in their pre-processing for application of the HDT approach for ordering the sampling points according to the pre-selected set of water quality parameters describing the ecological status of the river in most reliable way. The ordering was obtained by the use of water quality norms according to the Bulgarian environmental legislation in order to assess in detail the water quality of the whole river system. Thus, the combination between two chemometric data treatment strategies and the national water quality norms made it possible to achieve a complete surface water quality expertise.

1. Introduction

The assessment of surface water quality is an extremely important environmental issue. It is usually based on direct comparison of the water parameters measured values with threshold limits defined empirically. A much more reliable approach for classification, modeling and interpretation of data obtained from monitoring studies of surface water appears to be chemometrics using intelligent data analysis and data mining. Only multivariate statistical methods can describe the complex relationships in an ecosystem. In most published studies to date [1–12] concerning the problem of assessing the quality of surface water traditional chemometric methods are used: cluster analysis, principal components analysis, etc.

This study intends to combine the advantages of two integrated approaches to assess the water quality, such as self-organizing maps (SOM) and the Hasse diagrams technique (HDT). It has to be noted that unlike the classical chemometric methods as cluster analysis and principal components analysis, the SOM approach offered by Astel et al. [13], makes it possible to reveal specific features of the sampling sites within the monitoring net along a big river catchment and to detect additional hidden sources of pollution along the catchment.

On the other hand, the application of HDT for visualization of partial order relations between different environmental objects such as lake sediments [14–16], surface waters [17] or scenarios [18,19], offers an excellent expertise in assessing the anthropogenic influences since it makes possible to incorporate legislation requirements (principally univariate) in a multivariate mode of water quality estimation [17].

It is the aim of the present study to offer chemometric expertise in assessing the water quality along the stream of Mesta River on Bulgarian territory using routine monitoring data for a long period of observation, their self-organizing map clustering, Hasse diagram technique for partial ordering of different group of sampling locations, and water quality categorization using the local environmental legislation rules.

2. Experimental

2.1. Sampling area

Mesta (Greek Νέστος, Nestos) is a river that flows through southwestern Bulgaria and the territory of Greece. In ancient times the river was known by the names Nestos, Nesos and Mestos formed by the merger of the Cherna Mesta River and Byala Mesta River that spring from the Rila Mountain (Fig. 1). The two rivers merge northeast of city Yakoruda. Length of the river in its Bulgarian part is 126 km, on Greek territory, about 130 km. The average altitude of the Bulgarian part of the basin is 1318 m—the highest valley in Bulgaria. Catchment basin of the territory of Bulgaria covers an area of 2767 km^2, located between Rila to the north, Pirin to the west and Rhodope Mountain to the east. The south border, concurs with the state border with Greece,

* Corresponding author at: Faculty of Chemistry and Pharmacy, University of Sofia, 1, J. Bourchier Blvd., 1164 Sofia, Bulgaria. Tel.: +359 2 8161426; fax: +359 2 9625 438.
E-mail address: voyslavov@abv.bg (T. Voyslavov).

0169-7439/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2012.05.011

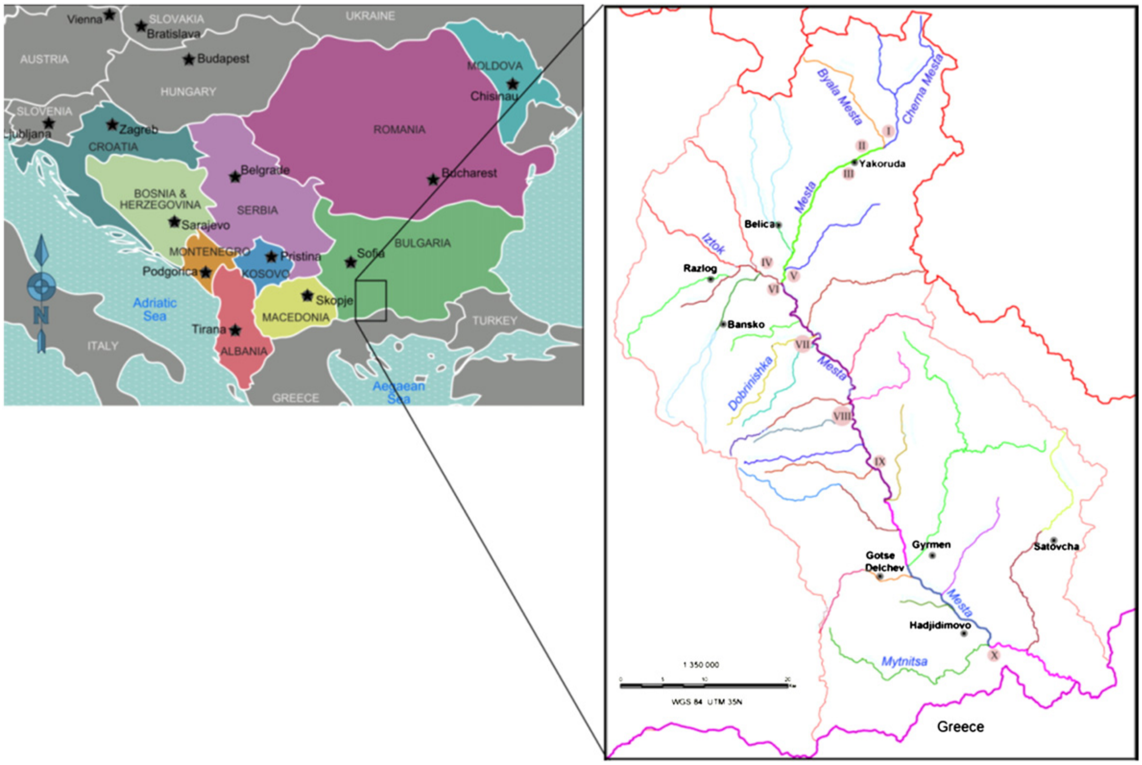

Fig. 1. Location of the sampling points on the Mesta River (Bulgaria).

is a valley of Western Thrace. Mesta River flows into the Aegean Sea near the island of Thasos.

Mesta has about 25 tributaries as the major ones on Bulgarian territory are Iztok, Bistrica, Kanina, Mytnitsa, Razlojka, Zlataritsa, Dobrinishka, etc.

There are a few larger towns in Mesta's catchment basin: Gotse Delchev (21000 inhabitants); Razlog (13000), Bansko (8500); Yakoruda (5700); etc. Total population in Mesta's valley is about 100000. In Bansko and Razlog there are sewage systems, but in both towns there were no treatment plants, so all waste waters were discharged without treatment in the near flowing rivers which are tributaries of the Mesta River.

Mesta monitoring data could be conditionally divided into two major sampling seasons: winter from October to March (high-water) and summer from April to September (shallow).

Water samples were collected between 1990 and 2009. The monitoring system covers ten sample sites where water quality is tested regularly on a daily, weekly or monthly basis. The data set used for the chemometric exploration consists of more than 6000 measurements on the Mesta River. The sites chosen almost completely cover the length of the river from its source to the Greek border. Due to missing data from the monitoring events the number of actual measurements used for statistical analysis is less than 6000.

The coding of the sample objects included the number of the sample sites (with Roman numerals from I till X), then 0 or 1 for two seasons of Mesta River (0 for shallow and 1 for high-water) and the year of sampling (two digits for each, e.g. 90, 91, 01, etc.). In the data set the objects were coded as III_1_95, IV_0_07, X_1_03, etc. The sample sites are presented in Table 1. Seasonal averages for the quality parameters were used.

The water quality parameters involved were dissolved oxygen (Diss O_2) [mg O_2 L^{-1}], saturated oxygen (Satur O_2) [%], oxidation ability (OXIS) [mg O_2 L^{-1}], biological oxygen demand for five days (BOD_5) [mg O_2 L^{-1}], dissolved matter (Diss sol) [mg L^{-1}], non-dissolved matter (Susp) [mg L^{-1}], ammonium (NH_4) [mg N–NH_4 L^{-1}], nitrate (NO_3^-) [mg N–NO_3 L^{-1}], nitrite (NO_2^-) [mg N–NO_2 L^{-1}], phosphate (PO_4^{3-}) [mg L^{-1}] and conductivity (COND) [μS]. The chemical analyses were performed according to standard analytical methods as routinely applied in the monitoring network's laboratories. Analytical methods such as potentiometry, titrimetry, gravimetry and spectrophotometry are the standard methods used in surface water quality analysis, especially for major indicators like those mentioned above. Sample preparation and sample measurements are described in detail elsewhere [20].

The surface water quality norms were extracted from Directive 7 issued by the Bulgarian Ministry of Environment and Water [21], and they are presented for comparison and interpretation along with the basic statistics of the input data in Table 2.

2.2. Chemometrics

2.2.1. Self-organizing maps of Kohonen

The SOM, also known as Kohonen map [22] is a neural-network model for exploring and visualizing high dimensional patterns in

Table 1
The sample sites description.

Number of sample sites	Short note for the sample sites
I	Biala Mesta River before the inlet of Cherna Mesta River
II	Mesta River before Yakoruda town
III	Mesta River after Yakoruda town
IV	Mesta River before the inlet of Iztok River
V	Mesta River before the inlet of Razlojka River
VI	Mesta River after the inlet of Razlojka River
VII	Mesta River 2 km after the inlet of Dobrinishka River, near to Filipovo village
VIII	Mesta River near to Momina kula area
IX	Mesta River near to Bukovo village
X	Mesta River after the inlet of Mytnitsa River, near to Hadjidimovo village

Table 2
Basic statistics and surface water quality norms (in mg/L except "Satur O_2" and "COND") n = 147 (sampling situations).

Parameter	Min	Max	Mean	Standard deviation	Surface water quality norms		
					I	II	III
Diss O_2	5.60	13.00	9.54	1.33	6.00	4.00	2.00
Satur O_2 (in %)	57.80	117.60	91.56	8.68	75.00	40.00	20.00
COND (in µS)	19	337	174	63	700	1300	1600
Susp	2	131	36	21	30	50	100
Diss sol	9	230	123	42	700	1000	1500
NH_4^+	0.002	2.000	0.269	0.400	0.100	2.000	5.000
NO_2^-	0.001	0.047	0.008	0.010	0.002	0.040	0.060
NO_3^-	0.01	11.28	1.34	2.26	5.00	10.00	20.00
PO_4^{3-}	0.002	1.200	0.268	0.260	0.200	1.000	2.000
BOD_5	0.05	7.67	2.26	1.42	5.00	15.00	25.00
OXIS	0.10	18.20	4.82	3.31	10.00	30.00	40.00

high dimensional data sets. SOM can be considered as non-linear mapping technique which identifies clusters in an unsupervised way in the data sets without rigid assumptions of linearity or normality of traditional statistical techniques. Unlike other neural networks SOM approach does not need any target output and it is an unsupervised pattern cognition method like cluster and principal component analysis. The SOM consists of two layers: the input layer which groups data according to their similarity, and the output layer of neurons arranged as a two-dimensional map. The grouping is performed as both measured objects and neurons (nodes) are presented like *n*-dimensional vectors, where *n* is the number of measured parameters (variables). After normalization the objects from the multidimensional data set are asserted to each individual node.

The Kohonen training clustering algorithm is based on "winner-take-all" rule, where the "winner" is the node whose vector most closely matches to the input sample vector. The winning node adjusts its vector weights to match the input vector, whereas the vectors in the nodes surrounding the "winner" are modified to look less like input vector. The projection of the multi-dimensional data set to two-dimensional map is usually done by calculation of the ratio between two largest input data eigenvalues according to the size and unit numbers of the map thus preserving the distances between objects in the initial data space. The trained map presents graphically the clustering of the objects and the variables distribution. The variables distribution is visualized in the form of easily interpretable 2D planes. Plane ordering is done by correlation between the planes and reveals the relationships between variables and could be used as variable selection procedure. This simultaneous presentation makes SOM an appropriate tool for revealing and identification of "hidden" different patterns among objects and measured parameters. SOM gives also a U-matrix plane, where distances between nodes are visualized. Additionally, the node vectors of the respective objects could be used for further statistical data treatment. In this study all calculations concerning SOM were performed by a free SOM Toolbox 2.0 [23].

2.2.2. Hasse Diagram Technique

Hasse diagrams visualize partial order relations between objects described by certain number of variables. HDT is well described elsewhere [14,24] and here only brief description concerning present study will be presented.

In HDT the ranking of objects (set of sampling stations during the whole sampling period, *E*) is done with respect to all variables (water quality parameters), which is called the "information basis" (IB). The processed data matrix **Q** (N×R) contains *N* objects and *R* variables. The entry $y_r(i)$ of **Q** is the numerical value of the *r-th* variable of the *i-th* object. The two objects *s* and *t* are comparable if:

$$s, t \in E; s \le t \Leftrightarrow y(s) \le y(t)$$

$$y(s) \le y(t) \Leftrightarrow y_r(s) \le y_r(t) \text{for all } y_r \in IB$$

If there is at least one y_r for which $y_r(s) > y_r(t)$ then the objects *s* and *t* are incomparable. Partially ordered sets (posets) could be easily developed by Hasse matrix which collects relations between each pair of objects. Hasse matrix is a (N×N) antisymmetric matrix where for each pair of elements *s* and *t* the entry h_{st} is given:

$$h_{st} \begin{cases} +1 & if \;\; y_r(s) \ge y_r(t) \text{for all } y_r \in IB \\ -1 & if \;\; y_r(s) < y_r(t) \text{for all } y_r \in IB \\ 0 & \text{otherwise} \end{cases}$$

If there is no element "*a*" of *E*, for which: $s \le a \le t$, $a \ne s$, *t* and $s \ne t$, then *s* is covered by *t* or *t* covers *s*.

The order relations stored in Hasse matrix could be visualized by Hasse diagram which is constructed as follows:

1. Each element is represented by a circle with an appropriate identifier. The circle could represent also equivalence class. Equivalent elements are different elements that have the same values for all variables included in IB.
2. If a cover relation holds, then a line between the corresponding object-pair is drawn and elements are comparable.
3. If $s \le t$ then *s* is drawn below (over) *t* and all relation lines should be done in the same downwards (upwards) direction. In present study the objects near the upper part of the Hasse diagram indicate objects that are the "affected" ones according to the criteria used for their ranking.
4. If $s \le t$ and $t \le z$ then $s \le z$ according to the transitivity rule. However line between *s* and *z* is not drawn. The relation is presented by lines between *s* and *t* and *t* and *z*.
5. If either $s \le t$ by some attributes or $t \le s$ by some others then *s* and *t* are not connected by a line and the objects are incomparable.

The objects not "covered" by other objects are called *maximal objects*. Those objects, which do not cover other objects, are called *minimal objects*. In some diagrams *isolated objects* also exist which are positioned as high as possible and are incomparable to all other objects. A *chain* is a set of comparable objects, therefore levels can be defined as the longest chain within the diagram. An anti-chain is a set of mutually incomparable objects, located at one and the same level, but not each anti-chain is a level. The height of the diagram is the longest chain and longest anti-chain is its width.

Posets have a structure, such as chains and anti-chains. Attributes have an impact on this structure. The quantification of this impact is the task of the sensitivity analysis. The sensitivity analysis of Hasse diagram towards variables describing objects could be done by the distances between posets induced by the full set of *R* variables and all possible subsets of *R*-1 variables. The importance of each variable for partial ordering is measured by the distance between poset induced by all *R* variables and the poset induced when the respective variable is omitted. Higher distance between diagrams corresponds to a higher importance of omitted variable for partial ordering.

All calculations concerning HDT were performed by the software package WHASSE [14,24] and DART [25].

The selection of relevant variables and proper grouping of objects are very important for meaningful Hasse diagram analysis and reliable evaluation. This problem is typical for multi-criteria evaluation but, usually, in partial order it leads to too many incomparable objects and to very complex Hasse diagrams. For reducing the complexity of data set and "noisy" differences between numerical parameter values, three groups of preprocessing techniques could be used. The first group includes techniques like Principal Component Analysis which reduces the number of variables [26], although PCA is in general not a method favoring a partially ordered set analysis, because of the orthogonality of the components the number of incomparabilities is enhanced. The second group of methods includes Cluster Analysis [26] and the SOM [16] which could be used for reducing the number of objects by introducing "artificial" representatives like cluster centroids or node output vectors. The SOM approach could treat both variables and objects [16]. The third group of methods like rounding values and bins partition are used for removing of numerical "noise" in the variables.

Thus, the purpose of the pre-processing techniques is to produce an "improved" data set for ranking exploration without loss of any relevant information. In the present study SOM and bins partition as preprocessing techniques are used.

3. Results and discussion

The data set used for this study consists of 147 objects as each one is described by 11 variables derived on seasonal basis (Table 2).

In Fig. 2 the U-matrix and all variable planes for the input data set are shown. Using colour scale the distribution of each variable on the SOM and the distances between nodes in U-matrix plane could be easily found. For example, the objects with high values for NO_3^- are located in the down left part of the SOM plane, while objects with high Susp are placed in the downright part of the plane.

The grouping of the variable planes (Fig. 3) shows four well-defined groups of correlated variables and some variables with specific location. The first group includes the water quality parameters PO_4^{3-} and NO_2^-. This fact shows the similar distributions of these parameters among the objects. The second well defined group is formed by NO_3^-, OXIS and BOD_5. The third group comprises the water parameters Diss O_2 and Satur O_2. The fourth group is formed by COND and Diss sol. The non-conventional positions of Susp and NH_4^+ could be explained by their ability to describe various complex pollutants and their transformations.

Using the ordering planes, a proper selection of surface water quality parameters could be done. Each well defined group could be selectively presented by one of its members. Thus, PO_4^{3-} and NO_3^- water quality parameters were selected to represent the first and the second group, respectively. The members of the fourth group (COND and Diss sol) were not selected since they do not possess sufficient discriminating ability since their maximum values are significantly lower than the surface water quality norms for drinking

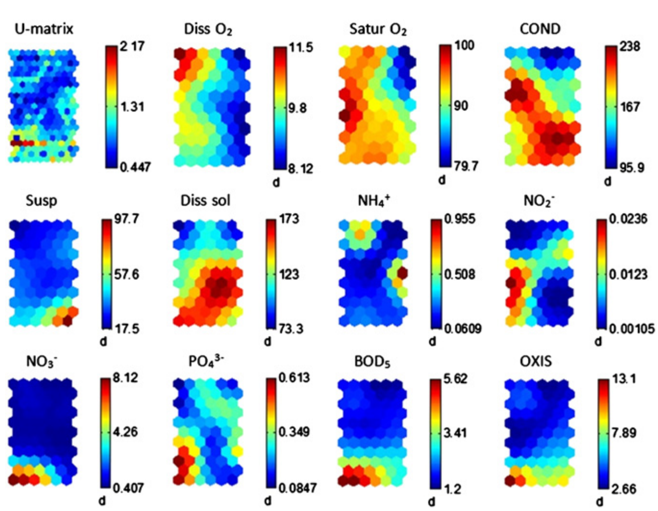

Fig. 2. U-matrix and variable planes for the input data.

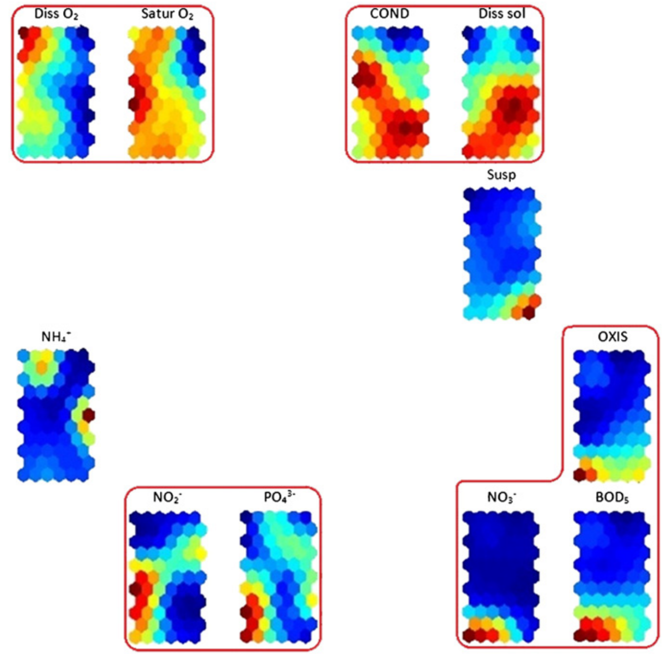

Fig. 3. Ordering of component planes.

waters (quality norm I). BOD_5 is chosen to represent all parameters with oxidation ability (OXIS, Diss O_2, Satur O_2 and BOD_5) although they are in different groups. It is known that Diss O_2 and Satur O_2 are temperature dependent and, thus, introduce seasonal drift. On the other hand, BOD_5 is more sensitive than OXIS to anthropogenic influences. The last two selected variables Susp and NH_4^+ are those with specific position as they reflect the complex nature of possible pollution sources. The selected water quality parameters could be more reliably and accurately analytically determined and are directly related to specific anthropogenic influences along the river catchment.

The new data set of 147 objects described by the five selected water quality parameters was further subjected to chemometric treatment by SOM approach. The ordering of the parameter planes and hit diagram of objects are presented in Fig. 4.

The hit diagram presents 147 objects (sampling sites and period of sampling) projected on 2D map with dimensionality 6×11. The objects are grouped in 53 plane units (nodes). The node numbering starts from the upper left part of the diagram and goes down to lower left part (nodes 1–11), then the same configuration is repeated for the next parallel sets of nodes (up, left to down left: 12–22; 23–33; 34–44; 45–55 and 56–66). The last node 66 is located at the right down part of the hit diagram.

Each populated node could be used as a representative element of an equivalence class, which includes the matching sampling situations and is presented by the respective node vector.

Using the data about the surface water quality norms in Bulgaria (Table 2) the coordinates of populated nodes (the new values of five selected parameters) will be binned in regular intervals (bins). It has to be mentioned that norm I reflects drinking water quality requirements; norm II reflects surface water quality for recreation purposes; and norm III reflects surface water quality for irrigation and industrial purposes. Values below norm I will be set as 0, values between norms I and II will be set as 1, values between norms II and III will be set as 2, and values higher than norm III will be set as 3. This operation makes the data set more homogeneous leading to decrease of the number of incomparable objects by reducing the non-relevant information (noise) in the water quality parameters. Thus, a more useful and interpretable Hasse diagram analysis could be performed. Pre-processing of

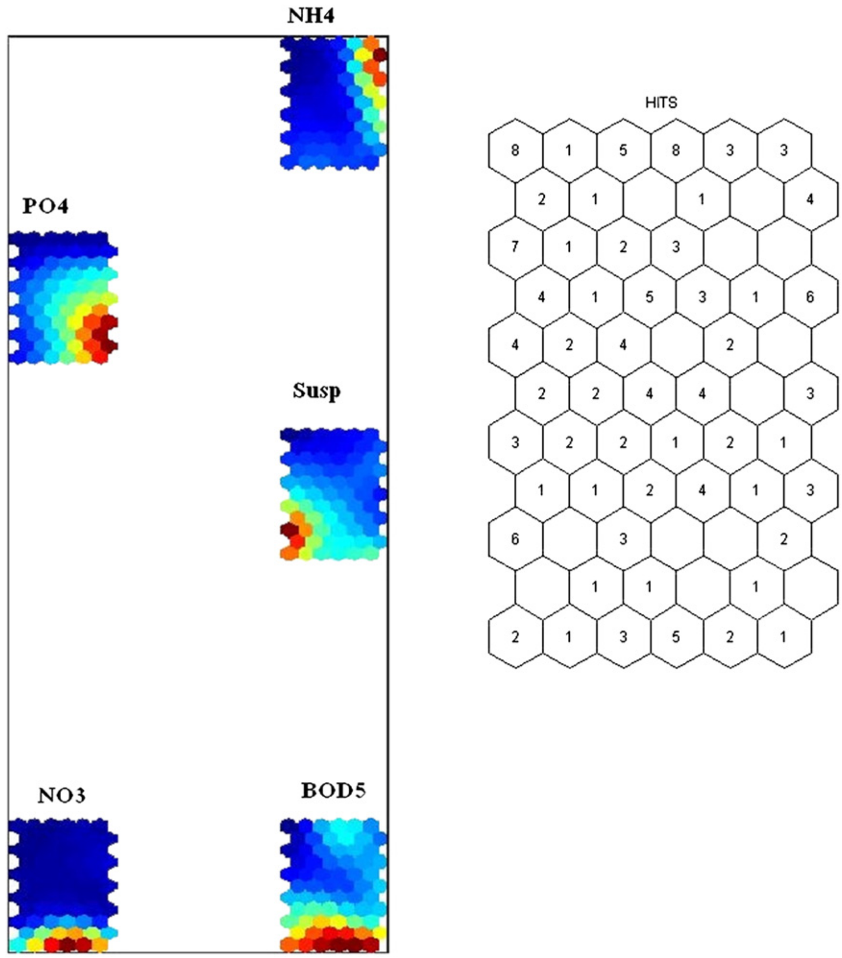

Fig. 4. The new parameter set ordering and respective hit diagram.

variables (bins partition) leads to formation of eleven equivalent classes where the nodes presenting objects have the same numerical values for all variables (Table 3).

The Hasse diagram presented for these 11 classes and 5 water quality parameters is shown in Fig. 5. The order of the parameters in the diagram is as follows: Susp, NH_4^+, NO_3^-, PO_4^{3-} and BOD_5.

The Hasse diagram (Fig. 5) has six levels, three maximal classes (19; 22 and 33) and one minimal (1). It is not surprising that equivalent class 1 collects the majority of sampling situations from remote sample points I and II. The equivalent class 22 consists of only one object—VII_0_94. The difference between class 22 and class 7 is only with respect to the content of NO_3^-. Both classes include mainly shallow water seasons from 1994 to 1997. It could be concluded that the increased value of NO_3^- for object VII_0_94 is due to some accidental local event with negligible environmental impact on the Mesta River flow. The similar situation is observed in equivalent class 19 where high values of PO_4^{3-} are detected in the two objects (VII_0_95 and X_1_05). So equivalent classes 19 and 22 could be ignored from further analysis because of their accidental nature.

A new Hasse diagram for nine classes and the same number of water quality parameters is shown in Fig. 6. The order of variables is the same as in the previous diagram.

The Hasse diagram has again six levels, two maximal classes (7 and 33) and only one minimal (1). It could be concluded that the two maximal classes determine two directed subgraphs in the Hasse diagram. The classes related to class 33 (33-32-17-26-34-1 and 33-32-17-6-4-1) form a directed subgraph initiated by the high values of nitrate, phosphate and BOD_5. The reason for formation of the second directed subgraph (7-4-1 and 7-34-1) is the high level of non-dissolved matter values (Susp) in class 7.

For our next analysis we will use the data from Hadjidimovo station. The station near Hadjidimovo (X) has worked for about twenty years, and it is the last sample point in the Bulgarian Mesta catchment region (Table 4).

Since the beginning of the monitoring period for which data are available, there are stable concentration values of all five considered parameters, as seen in class 33 of the top level of the diagram. In the next two years BOD_5 values decrease, and objects fall in class 32. This is an indication for a rapid clean up of the river from readily biodegradable pollutants. Next years (1993–1995) are characterized by decreasing of nitrate values and falling to class 17. Between 1995 and 1997 the objects from Hadjidimovo were dominantly placed in the second subgraph of the diagram in class 7, which is a sign for another kind of contamination, including higher Susp values. Such kind of influence also could be observed in the shallow seasons of 1992 and 1994. In the next two years the objects belong to class 17 again, because of decreased Susp concentrations. During 2000 the objects from Hadjidimovo belong to class 34, which is stable with respect to ammonia levels and in 2001 reached the minimal level in the diagram corresponding to waters of I quality norm. Since 2002 till 2008 the high-water seasons of Hadjidimovo are dominantly “lodged” in class 17. The respective shallow water seasons and the end of the monitoring period (2009) are located in classes 6 and 26 where the presence of elevated phosphate concentrations should be mentioned. The other nine sampling stations, working temporary during monitoring period, follow the behavior of Hadjidimovo.

Table 3
Equivalent classes representing the sampling stations with their coded surface water quality indexes.

Equivalent class *	Code of the sample object	Parameters				
		Susp	NH_4^+	NO_3^-	PO_4^{3-}	BOD_5
1 (1, 2, 3, 12, 13, 14, 23, 25)	I (1_05; 1_06; 0_07; 1_07); II (1_93; 1_94; 1_95; 1_96; 1_98; 1_99; 0_00; 1_00; 0_01; 1_01; 0_02; 1_03; 0_04; 1_05; 1_08; 0_98); V (1_97; 0_01); VIII (0_00; 0_01); X (1_98; 0_01; 1_01)	0	0	0	0	0
4 (4, 5)	II (1_02; 1_04); V_1_03; VIII (0_03; 1_03; 0_04; 0_05); X_0_04	1	0	0	0	0
34 (34, 35, 45, 56, 57)	V (0_93; 1_93; 0_95; 1_95; 1_96; 0_97; 0_99; 1_00; 1_01); VI (0_93; 1_94); VII(1_96; 1_98); VIII_1_01; X (0_93; 1_96; 0_97; 0_00; 1_00);	0	1	0	0	0
6 (6, 15, 16, 27)	II_0_03; V (0_03; 0_04; 0_05); VIII_0_02; X (0_03; 1_03; 0_07; 1_09)	1	0	0	1	0
26 (26, 36, 37, 61, 62, 63)	II (0_98; 0_99); V (0_94; 0_98; 1_98; 1_99; 0_00; 0_02; 1_08); VII_0_97; VIII (0_98; 1_00; 0_07; 1_07; 1_08); X (0_02; 0_08; 0_09)	0	1	0	1	0
17 (17, 18, 28, 29, 30, 31, 39, 40, 41, 48, 49, 51, 52, 54, 59, 64)	I (0_05; 0_06); III (1_08; 0_08); IV (0_09; 1_09); V (1_94; 1_02; 1_04; 1_05; 0_06; 1_06; 0_07; 1_07; 0_08); VII (1_92; 0_93); VIII (1_98; 0_99; 1_99; 1_02; 1_04; 1_05; 0_06; 1_06); IX (1_09; 0_09); X (1_93; 1_94; 0_95; 0_98; 0_99; 1_99; 1_02; 1_04; 0_05; 1_06; 1_07; 1_08)	1	1	0	1	0
32 (32)	**X**_1_92	1	1	1	1	0
33 (33, 44, 55, 66)	VII (0_90; 1_90; 0_91; 1_91; 0_92; 1_93; 1_94); X (0_90; 1_90; 0_91; 1_91)	1	1	1	1	1
7 (7, 8, 9, 11)	III_0_09; V_0_96; VII (1_95; 0_96; 1_97); VIII_0_08; X (0_92; 0_94; 1_95; 0_96; 1_97; 0_06)	2	1	0	0	0
19 (19, 21)	VII_0_95; X_1_05	2	1	0	1	0
22 (22)	VII_0_94	2	1	1	0	0

* In parentheses SOM node numbers are given.

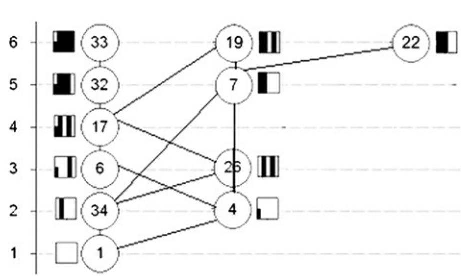

Fig. 5. Hasse diagram for eleven classes (parameters order from left to right: Susp, NH_4^+, NO_3^-, PO_4^{3-} and BOD_5).

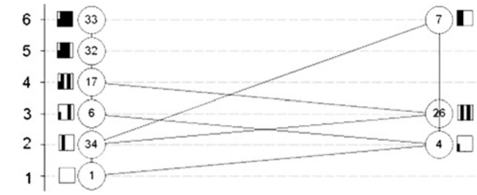

Fig. 6. Hasse diagram for nine classes (parameters order from left to right: Susp, NH_4^+, NO_3^-, PO_4^{3-} and BOD_5).

Usually the higher values of phosphate and suspended particulate matter are markers for pollution due to agricultural activity [27], which is the main pollutant source with an important impact until the mid-1990s of the last century (classes 33, 32 and 17). The presence of phosphate and ammonia in the class 26 dated from a later shallow seasons shows a second pattern of contamination. Taking into account the location of the basin of the river, we may conclude that the new kind of contamination is due to the urban development caused by touristic activity (increase number of tourists, hotels, restaurants, laundries) in the skiing resort town of Bansko and neighbouring towns in the Mesta River valley. Both pollutant patterns overlap in the years, so that the concentration of phosphate remains unchanged for an extended period of time. Objects of the recent years of the monitoring belong to the classes located on level 3 and 4 on the Hasse diagram, which is a sign of recovery of the agricultural activity in the region. Lack of objects in the “touristic activity” equivalence classes by the end of the monitoring period may be interpreted by the construction of waste-water treatment facilities in some of the towns along the river flow.

The calculated sensitivities of water quality parameters for Hasse diagram are presented in Table 5. The most important parameter for current partial ordering is Susp followed by NH_4^+ and PO_4^{3-}. These results confirm the importance of above mentioned three parameters for surface water quality and reveal agricultural and touristic anthropogenic influences.

4. Conclusion

The present study offers a reliable scheme for assessment of the water quality along the Mesta River flow on Bulgarian territory. The integration of the multivariate statistical methods self-organizing maps of Kohonen (SOM) and Hasse diagram technique (HDT) for partial ordering contributes to the realization of the following features of this environmental study:

- Careful selection of variables to be included in the partial ordering (using the grouping of the water parameters in the component planes by SOM method);
- Distribution of the sampling events (site location and period of sampling) on the SOM hit diagrams and construction of equivalent

Table 4
Hadjidimovo (X) sampling situation locations in the equivalent classes of the Hasse diagram.

Equivalent class	Hadjidimovo (X) sampling situations
1	1_98; 0_01; 1_01
4	0_04
34	0_93; 1_96; 0_97; 0_00; 1_00
6	0_03; 1_03; 0_07; 1_09
26	0_02; 0_08; 0_09
17	1_93; 1_94; 0_95; 0_98; 0_99; 1_99; 1_02; 1_04; 0_05; 1_06; 1_07; 1_08
32	1_92
33	0_90; 1_90; 0_91; 1_91
7	0_92; 0_94; 1_95; 0_96; 1_97; 0_06

Table 5
Sensitivity analysis results.

Water quality parameters	Sensitivity
Susp	10
NH_4^+	5
NO_3^-	1
PO_4^{3-}	5
BOD_5	1

classes depending on the water quality norm values as required by the national legislation (using only the pre-selected variables proven to be the most representative for the water quality);

- Construction and reconstruction of Hasse diagrams to explain the relationships between the objects belonging to various equivalent classes and the variables describing them;
- Specific interpretation of the spatial and temporal changes of the water quality for each one of the sampling sites.

Acknowledgement

The authors would like to express their sincere gratitude to the National Science Fund (Project DO-02-352) for the financial help. One of the authors (S.T.) acknowledges the support of National Science Fund of Bulgaria (grant No. DTK02/58).

References

[1] A. Astel, G. Głosińska, T. Sobczyński, L. Boszke, V. Simeonov, J. Siepak, Chemometrics in assessment of sustainable development rule implementation, CEJCh 4 (3) (2006) 543–564.

[2] N.M. Mattikalli, Time series analysis of historical surface water quality data of the River Glen Catchment, UK, Journal of Environmental Management 46 (2) (1996) 149–172.

[3] C. Cun, R. Vilagines, Time series analysis on chlorides, nitrates, ammonium and dissolved oxygen concentration in the Seine river near Paris, Science of the Total Environment 208 (1–2) (1997) 59–69.

[4] D.A. Wunderlin, M.P. Diaz, M.V. Ame, S.F. Pesce, A.C. Hued, Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía River Basin (Córdoba–Argentina), Water Research 35 (12) (2001) 2881–2894.

[5] R. Götz, B. Steiner, P. Friesel, R. Klaus, F. Walkow, V. Maaβ, H. Reincke, B. Stachel, Dioxin (PCDD/F) in the river Elbe—investigations of their origin by multivariate statistical methods, Chemosphere 37 (9–12) (1998) 1987–2002.

[6] V. Simeonov, J.W. Einax, I. Stanimirova, I. Kraft, Environmetric modelling and interpretation of river water monitoring data, Analytical and Bioanalytical Chemistry 374 (5) (2002) 898–905.

[7] C. Mendiguchía, C. Moreno, D.M. Galindo-Riaño, M. García-Vargas, Using chemometric tools to assess anthropogenic effects in river water. A case study: Guadalquivir (Spain), Analytica Chimica Acta 515 (1) (2004) 143–149.

[8] S. Stefanov, V. Simeonov, S. Tsakovski, Chemometrical analysis of waste water monitoring data from Yantra river basin, Bulgaria, Toxicological and Environmental Chemistry 70 (3) (1999) 473–482.

[9] T. Kowalkowski, R. Zbytniewski, J. Szpejna, B. Buszewski, Application of chemometrics in river water classification, Water Research 40 (4) (2006) 744–752.

[10] S. Tsakovski, V. Simeonov, S. Stefanov, Time-series analysis of long-term water quality records from Yantra river basin, Bulgaria, Fresenius Environment Bulletin 8 (1999) 28–36.

[11] E. Marengo, M.C. Gennaro, D. Giacosa, C. Abrigo, G. Saini, M.T. Avignone, How chemometrics can helpfully assist in evaluating environmental data lagoon water, Analytica Chimica Acta 317 (1–3) (1995) 53–63.

[12] D. Brodnjak-Vončina, D. Dobčnik, M. Novič, J. Zupan, Chemometrics characterization of the quality of river water, Analytica Chimica Acta 462 (1) (2002) 87–100.

[13] A. Astel, S. Tsakovski, P. Barbieri, V. Simeonov, Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets, Water Research 41 (19) (2007) 4566–4578.

[14] R. Brüggemann, E. Halfon, G. Welzl, K. Voigt, C. Steinberg, Applying the concept of partially ordered sets on the ranking of near-shore sediments by a battery of tests, Journal of Chemical Information and Computer Science 41 (4) (2001) 918–925.

[15] H. Hollert, S. Heise, S. Pudenz, R. Brüggemann, W. Ahlf, T. Braunbeck, Application of a sediment quality triad and different statistical approaches (Hasse diagrams and fuzzy logic) for the comparative evaluation of small streams, Ecotoxicology 11 (5) (2002) 311–321.

[16] S. Tsakovski, V. Simeonov, Hasse diagram technique as exploratory tool in sediment pollution assessment, Journal of Chemometrics 25 (5) (2011) 254–261.

[17] S. Tsakovski, A. Astel, V. Simeonov, Assessment of the water quality of a river catchment by chemometric expertise, Journal of Chemometrics 24 (11–12) (2010) 694–702.

[18] U. Simon, R. Brüggemann, S. Pudenz, Aspects of decision support in water management – example Berlin and Potsdam (Germany) I – spatially differentiated evaluation, Water Research 38 (7) (2004) 1809–1816.

[19] U. Simon, R. Brüggemann, S. Pudenz, Aspects of decision support in water management – example Berlin and Potsdam (Germany) II – improvement of management strategies, Water Research 38 (19) (2004) 4085–4092.

[20] Bulgarian State Standards, EN and ISO, Sofia, 1985.

[21] Bulgarian Ministry of Environment and Water, State directive 7 (Surface water quality norms), 1997.

[22] T. Kohonen, Self-organization and Associative Memory, 78(9), Springer-Verlag, Berlin, 1984, pp. 1464–1480.

[23] J. Vesanto, SOM-based data visualization methods, Intelligent Data Analysis 3 (1999) 111–126.

[24] R. Brüggemann, K. Voigt, C. Bücherl, Theoretical base of the program “Hasse”, GSF—Bericht 20/95, Neuherberg, Software product which may be obtained after personal contact to Dr Brüggemann (e-mail: brg_home@web.de), 1995.

[25] DART (Decision Analysis by Ranking Techniques) v.2.05, http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools/DART(accessed 15 March 2012).

[26] M. Pavan, R. Todeschini, Scientific Data Ranking Methods: Theory and Applications, Elsevier, Amsterdam, 2008.

[27] D. Berryman, État de l'écosystème aquatique du bassin versant de la rivière Yamaska: faits saillants 2004–2006, Québec, ministère du Développement durable, de l'Environnement et des Parcs, Direction du suivi de l'état de l'environnement, 978-2-550-53592-8, 2008.