Review of the Self-Organizing Map (SOM) Approach in the Field of Environmental Engineering

B. Sowmiya, S. Amal Raj

***Abstract:** In environmental engineering field, the use of artificial neural networks (ANNs) has received steadily increasing interest over the last decade or so. In ANN, self-organizing map (SOM) is an unsupervised learning method to analyze, cluster, and model various types of large databases. There is, however, still a notable lack of comprehensive literature review for SOM along with training and data handling procedures, and potential applicability. As a result the present paper, first explains the basic structure and algorithm of self-organizing map (SOM) and secondly, to review published applications with special importance on environmental engineering related problems in order to assess how well SOM can be used to solve a particular problem. Finally, concluded that self-organizing map (SOM) is a hopeful technique suitable to investigate, model, and rule environmental related problems. However, in recent years, self-organizing map (SOM) has displayed a steady increase in the number of applications in environmental engineering related problems due to the robustness of the method.*

***Keywords:** Linear and non- linear process, Artificial Neural Network, Self Organizing Map, Environmental Engineering, Review.*

I. INTRODUCTION

Modeling of environmental related problems that are embedded with high complexity and non- linearity in both spatial and temporal scales is an important task for environmental engineers. In many cases, however, the lack of physical understanding of the complex processes involved creates problems to find efficient models. Over the last decades artificial neural network (ANN) is being widely applied in environmental engineering field for solving various problems. Its ability to learn highly complex interrelationships based on the provided data sets, along with less amount of data requirement, makes it a powerful modeling tool. Artificial Neural Networks are relatively crude electronic models based on the neural structure of the brain and are capable of modeling nonlinear statistical data where the complex relationships between inputs and outputs. The theory and application of ANN in environmental engineering was well explained by Berthold and Mibradt. The self-organizing map (SOM), also called Kohonen map or topology preserving feature map) is a type of ANN method which is capable of clustering, classification, estimation, prediction, and data mining (Alhoniemi et al., 1999; Vesanto and Alhoniemi, 2000; Kohonen, 2001)

Revised Version Manuscript Received on July 29, 2016.

B. Sowmiya, Research Scholar, Centre for Environmental Studies, Anna University, Kotturpuram, Chennai (Tamil Nadu). India.

Dr. S. Amal Raj, Professor, Centre for Environmental Studies, Anna University, Kotturpuram, Chennai (Tamil Nadu). India.

In a wide-spread range of disciplines regarding signal recognition, organization of large collections of data, process monitoring and analysis, and modelling as well as water resources problems. Typical for SOM, the desired solutions or targets are not given and the network intelligently learns to cluster the data by recognizing different patterns.

Despite the rather broad existing literature about ANN methods, in particular feed-forward MLPs (i.e., Maier and Dandy, 2000; ASCE, 2000a,b; Dawson and Wilby, 2001), there is a notable lack of comprehensive literature review on the efficiency of unsupervised learning techniques. Therefore the main objective of this paper is to explain the SOM algorithm and to review the successes or failures of published applications with main emphasis on water resources and related disciplines. The paper is methodically divided into two main parts. First, SOM methods are explained along with a presentation of their structural differences. Secondly, published applications of the SOM method in environmental engineering problems and related disciplines are reviewed and evaluated. Finally the paper was concluded by giving a way of approaching a problem for the application of SOMs in environmental engineering.

II. SELF- ORGANISING MAPS

The Self-Organizing Map is one of the most popular neural network models. It belongs to the category of competitive learning networks. The Self-Organizing Map is based on unsupervised learning, which means that no human intervention is needed during the learning and that little need to be known about the characteristics of the input data. We could, for example, use the SOM for clustering data without knowing the class memberships of the input data. The SOM can be used to detect features inherent to the problem and thus has also been called SOFM, the Self-Organizing Feature Map. The Self-Organizing Map was developed by professor Kohonen.

A SOM provides a topology preserving mapping from high dimensional space to map units or neurons which form a two dimensional lattice. Topology preserving mapping connotes that the mapping conserves the relative distance between all the points i.e. points which were initially in the vicinity of each other are mapped to nearby map units in SOM. The main goal of the SOM is to transform an incoming pattern of arbitrary dimensions into a one or two dimensional discrete map. Each output neuron is fully connected to all the source nodes in the input layer. This network represents a feedforward structure with a single computational layer consisting of neurons arranged in a one or two dimensional grid. The topology of the grid can be square, hexagonal, etc. In this paper we concentrate on a particular type of SOM called the Kohonen's Network. Such an SOM has feed forward structure with a single computational layer arranged in rows and columns. The basic structure of SOM is shown in figure 1.

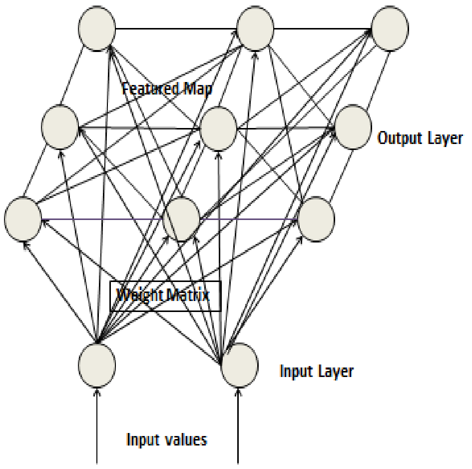

Fig 1: Structure of Self- Organizing Map

A. Components of Self Organization

The self-organization process involves four major components:

a) *Initialization:* All the connection weights are initialized with small random values.

b) *Competition:* For each input pattern, the neurons compute their respective values of a discriminant function which provides the basis for competition. The particular neuron with the smallest value of the discriminant function is declared the winner.

c) *Cooperation:* The winning neuron determines the spatial location of a topological neighbourhood of excited neurons, thereby providing the basis for cooperation among neighbouring neurons.

d) *Adaptation:* The excited neurons decrease their individual values of the discriminantfunction in relation to the input pattern through suitable adjustment of the associated connection weights, such that the response of the winning neuron to the subsequent application of a similar input pattern is enhanced.

B. The stages of the SOM algorithm can be summarised as follows

The operation of the SOM algorithm is represented as a series of the following steps and shown in Figure 2.

a) Step 1. Initialization – Choose random values for the initial weight vectors wj.

b) Step 2. Sampling – Draw a sample training input vector x from the input space.

c) Step 3. Matching – Find the winning neuron I (x) with weight vector closest to input vector.

d) Step 4. Updating – Apply the weight update equation $\Delta w_{ji} = \eta(t)\, T_{j,I(x)}\, (t)\, (x_i - w_{ij})$.

e) Step 5. Continuation – Keep returning to step 2 until the feature map stops changing.

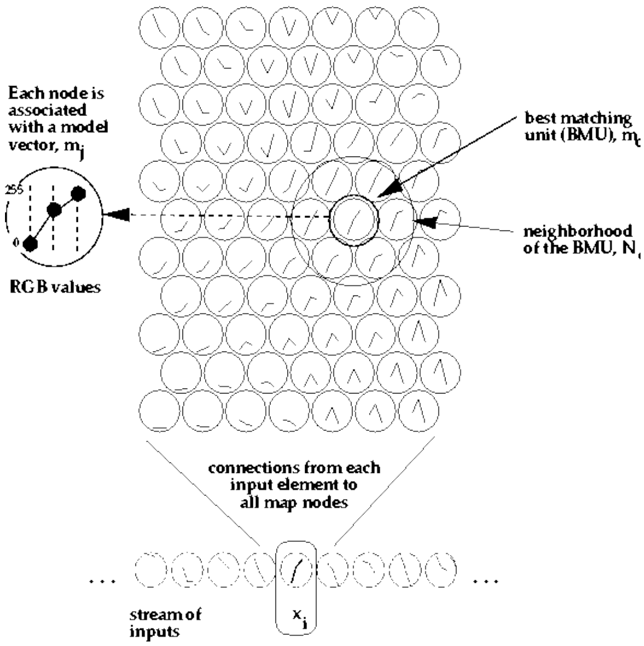

Figure 2: SOM algorithm structure

III. APPLICATION OF SOM IN ENVIRONMENTAL ENGINEERING

In earlier, we discussed briefly about SOM structure and its basic concepts along with the procedures that are required to apply SOM algorithm for a data set. Now, we review some of the SOM applications in environmental related problems.

A. SOM in wastewater

Heikkinen et al (2010), adapt SOM to optimize the activated sludge wastewater treatment process of an industrial pulp & paper mill. First, the process data was modeled using the SOM algorithm. Next, the reference vectors of the map were classified by K-means algorithm into clusters, which represented different states of the process. In final stage, the reference vectors of the map and the center vectors of the clusters were used for subtraction analysis to indicate differences of the process states. They extracted raw data from databases of the pulp mill. Variable selection was made by a process expert and the raw data of the selected variables was united into solid data matrix and the size of the data matrix was 29×1378 (29 variables in columns, 1378 rows).

They coded the data into 29 inputs for the SOM having 256 neurons in a 16 × 16 hexagonal arrangement was constructed. The linear initialization and batch training algorithms were used in training the map. A Gaussian function was used as the neighborhood function. The map was taught with 50 epochs and the initial neighborhood had the value of 6. The SOM Toolbox program (v. 2.0 beta) was used in the analysis under a Matlab-software platform. The K-means algorithm was applied to the clustering of the trained map, or more precisely, to the clustering of the reference vectors. By clustering the map the interactions can be detected more easily, and the clusters can then be treated as sub-models of the main model, which was formed by the SOM algorithm. After training and clustering, the desired reference vector elements of clustered neurons were visualised in a two-dimensional space to reveal the possible interactions between data variables. In the subtraction analysis, two vectors are subtracted from each other to indicate differences between these vectors. Using the centre vectors of the clusters, the following formula(1) was used

$$s=(C\alpha - C\beta) \qquad (1)$$

where *s* is the subtracted result vector, *Cα* is the centre vector of cluster *α*, and *Cβ* is the centre vector of cluster *β*. This method can be used for reveal differences in process factors between the main process states. Then the events of the same cluster can be compared for the optimal reference vector using the following formula:(2)

$$s= (t\gamma - R\gamma) \qquad (2)$$

where *s* is the subtracted result vector, *tγ* signifies the reference vector of the most optimal neuron in cluster *γ* and *Rγ* is the reference vector in cluster *γ*. The map and the clusters with centre vectors and the map and the clusters with short descriptions are represented in seperate figures. Therefore they concluded that, by using the selected method, interesting relations were found quite fast and easily,and the applied SOM-based neural network method is an efficient and fruitful way to model data acquired from the sludge treatment.

B. SOM in Water quality analysis

Bowden et al. (2002) presented two methods based on a genetic algorithm and an SOM which were able to divide the available data into subsets with common statistical properties. These were used to develop back-propagation ANN models for salinity forecasting in the River Murray at Murray Bridge, South Australia, 14- days in advance. They compared the performance of the obtained models with a model that was developed by means of arbitrary division of data. They found that the models developed using the presented techniques outperformed the conventional one. It was found that the SOM also could be used as a tool to find out why the ANN models can give poor results in some parts of the time series. To this end, they utilized an SOM for clustering the data. An SOM with 10 _ 10 neurons in the Kohonen layer was used so that the input patterns were clustered into 100 clusters, where 49 consisted of three or more patterns. From each of these clusters three data patterns were sampled, one each for training, testing, and validation. For the 51 clusters containing less than three patterns, they used the sample record in the cluster with only one record as training and in the case of clusters with two records, one was used for training and the other for testing. Another issue that also needs great attention in the development of ANN models is the determination of significant input variables. This is due to the fact that the presentation of all potential input variables to the ANN and relying on the network to identify the most critical ones may create problems.

Garcia et al (2004) used SOM for wastewater treatment plant performance monitoring. In this study they used the SOM toolbox (Vesanto et al., 1999) version 2.0 beta (10/10/00) developed at the HUT (Helsinki University of Technology). In a first stage, a selection of the most significant variables was done and in the second stage data were normalized to a zero mean value and a variance σ = 1 to bring all the variables in the same ranges. After normalising, Network was trained with these variables using batch training algorithm, which is a version of the SOM algorithm. The learning rate α (t) is not used in the batch algorithm. After training, the SOM is expected to capture the geometry of the process data, allowing it to be represented in 2D representations. Then the Visualization of results was done in several ways such as (i) inter neuron distance matrix (U matrix) (ii) component planes (iii) best clustering and (iv) condition achievement maps. The network structure was a 2D-lattice of 32×23 hexagonal units because the training data set is composed of 21,095 samples. Finally concluded that the combination of the component planes obtained by means of the SOM and the best clustering structure is a useful tool to visualize the data structure in order to label each state of the input data space onto the 2D lattice that allows the different features of the process to be monitored. Plant requirements were represented by means of the condition achievement maps and it shows the output space under several conditions of the variable values.

Bowden et al., (2005a, b) pointed out, there are several disadvantages with this approach, including the increase of computational complexity and memory requirements, difficulty in learning, misconvergence and poor model performance, increase of the complexity of the model and consequently, a difficulty in understanding the model as well as increasing noise due to inclusion of spurious input variables. Consequently, there is a need to introduce an efficient input determination method to select a more parsimonious model. This issue was addressed by Bowden et al. (2005a, b) who presented two methodologies, one based on the partial mutual information (PMI) algorithm and the other based on an SOM combined with a hybrid genetic algorithm and a general regression neural network (SOM-GAGRNN). The first uses a partial measure of the mutual information criterion in order to determine the inputs that have highly significant relationship with the output variable of the system being modeled and the second uses an SOM to cluster the input variables into groups of similar inputs. Then they selected one input variable from each cluster so that the input with the smallest Euclidean distance to the cluster's weights was selected from each cluster. Thereby the dimensionality of the input space was reduced and the variables obtained were introduced to the GAGRN to determine what inputs that have significant relationship with the output variable of the system being modeled. The authors tested the proposed methods on several synthetic data sets and concluded that in terms of predictive performance, both methods were good while from the point of view of getting valuable information about the system under investigation the first was recommended. To verify the above approaches on real data, they were applied for input determination for an ANN model in forecasting salinity 14-days in advance in the River Murray at Murray Bridge, South Australia. The input variables for the ANN included daily salinity, flow, and river level data at 16 locations along the river. With 60 lags, there were, thus, 960

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

input variables. The proposed techniques were compared with two methods used in previous studies by Maier and Dandy (1996, 1997) for forecasting salinity. The authors found that the new techniques led to more parsimonious (in terms of number of inputs) ANN models and claimed that the models developed by means of the new techniques had higher generalization ability.

Agaudo et al., (2008) analysed the evolution of an enhanced biological phosphorous removal (EBPR) process which was affcted by zoogloeal overabundance using Kohonen self-organizing maps (SOMs) as well as PCA to uncover the relationships among the collected process variables and to classify and interpret process behaviour. A laboratory SBR was operated with four 6-h cycles per day under anaerobic/aerobic conditions for EBPR from wastewater. The synthetic wastewater used in this research consisted of two separated solutions prepared with tap water. One containing mineral compounds and the other acetate and NH_4Cl (pH adjusted to 7.5 with NaOH). During the experimental period (almost 3 months of SBR operation), samples were periodically withdrawn from the reactor and analysed off-line in a quality control laboratory for several parameters such as VFA (Volatile fatty acids concentration in the influent), NH_4 (Ammonium concentration in the influent), TSS (Total suspended solids Concentration in the reactor), VSS (Volatile suspended solids concentration in the reactor), %VSS (Percentage of volatile suspended solids in the reactor), Ps (Soluble phosphorus concentration in the reactor at the end of the anaerobic stage), Δ_{DPana} (Difference between phosphorus concentration in the reactor at the end and at the start of the anaerobic stage),Pefl mg/l (Phosphorus concentration in the effluent) NO_3efl mgl_1(Nitrate concentration in the effluent) SVI ml/g Sludge volume index YPO4 Ratio between phosphorus release and acetate uptake (in the anaerobic stage).Additionally, microscopic observation of the sludge was also performed to assess the microbial community composition. EBPR is a wastewater treatment process aimed at achieving low phosphorus effluent concentrations. This process is based on the ability of certain type of bacteria (PAOs) to store phosphorus in higher amount than just that corresponding to nutritional requirements. The SOM Toolbox developed at Helsinki University of Technology as MATLAB 6.5, and SIMCA-P 9.0 was used for this work. The data set consisted of 11 variables and 328 observations (or samples), corresponding to the whole experimental period including the start-up process.

In order to make it more suitable for the analysis they pre-processed the raw data before the application of the visualization techniques. As a result of the pre-processing, all variables have zero mean and unit variance. The pre-processed data set was used to train a SOM network. The number of neurons (M) of the SOM was obtained from the heuristic formula $M = 5\sqrt{N}$, where N is the number of observations (samples) in the data set. The map resulted in a map with total neurons of 91 [13×7]. The topology of the SOM consisted in a local hexagonal lattice structure and a global sheet shape. The SOM was trained initially using batch training algorithm and finally using fine tuning algorithm. From the results they concluded that the that the applied techniques are highly effective to compress multidimensional data sets and to extract relevant information from the P-removal process making the interpretation and diagnosis much easier and evident.

Liukkonen et al., (2011) used SOM for modeling fluidized bed combustion process and No_x emissions. The authors described that the self-organizing map (SOM) provides an efficient method for revealing the most characteristic features of input data, making it a powerful tool for discovering general phenomena and visualizing the behavior and evolution of a combustion process. They analysed the process data using a SOM and K-means clustering to create subsets representing the separate process states in the boiler. The raw data for this work were extracted from the database of a utility-scale CFB boiler, with a time resolution of 15 min. The size of the data matrix used here as an example is 10000 rows × 38 variables in the columns, covering 104 days of operation. The boiler used in the analysis was typically operated in two modes according to load peaks, using different settings for high and low loads. Finally the authors concluded that, self-organizing maps have proved useful and efficient for emission modeling, diagnosis, and process optimization with respect to circulating fluidized beds (Heikkinen et al., 2005, 2009; Liukkonen et al., 2009b, c). And also concluded that the SOM method together with trajectory analysis provides a descriptive and functional means of visualizing and studying short-term fluctuations and long-term evolution in a combustion process ans also added that it is a useful to examine fluctuations from one process state to another, because they provide additional information on the operational use of the boiler and its consequences, such as variations in the levels of harmful emissions into the air.

Voyslavov et al., (2012) adapt SOM network & Hasse diagram technique for assesing surface water quality of Mesta River on Bulgarian territory. They divided the sampling seasons into summer and winter and the samples were collected between the period from 1990 to 2009. Monitoring systems cover ten sampling sites where the quality of water is tested regularly on daily, weekly or monthly basis according to satandard analytical methods. Here all calculations concerning SOM were performed by a free SOM Toolbox 2.0. Hasse diagrams visualize partial order relations between objects described by certain number of variables. All calculations concerning HDT were performed by the software package WHASE and DART. To reduce the complexity of data set and noisy differences between numerical parameters values, there are some preprocessing techniques such as PCA that reduces the number of variables. (Tsakovsk et al., 2011). In second group they used cluster analysis and SOM which will reduces the number of objects by introducing “artificial” representatives like cluster centroids. The Som approach could treat both variables and objects. (Pavan et al., 2008). The third group of methods like rounding values and bin partition are used for removing of numerical noise in the variables.

Here, they used SOM and bin partition as preprocessing techniques. The dataset used for this study was 147 objects

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

as each one is described by 11 variables derived on second basis. Finally, they concluded that the SOM and HDT explain the relationships between the objects belonging to various classes and variables describing them. Those techniques also interpret the spatial and temporal changes of water quality of sampling sites.

They concluded from the results and subsequent process interpretation, two clusters were found to be formed by the observations from the start-up and sludge adaptation periods; two clusters were clearly associated to good process conditions (with high PAO population, low effluent phosphorus concentration and excellent settling properties) while other two clusters to undesirable process status. The latter resulted from the appearance of an excess of Z.ramigera in the system, which competed for the substrate with PAO_s, resulting in low YPO_4 values. Quick detection of this type of adverse process conditions allows fast process operator intervention in order to recover process performance, thus, avoiding further process deterioration. Therfore, important cost-savings could be achieved in full-scale EBPR systems since the overabundance of zoogleal colonies implies additional substrate consumption and the poor effluent characteristics could require an extra treatment for meeting the quality requirements imposed by the regulations.so that they concluded that these results clearly say that the applied techniques are highly effective to compress multidimensional data sets and to extract relevant information from the P-removal process making the interpretation and diagnosis much easier and evident.

Olawoyin et al., (2013) categorized water, soil and sediment quality in Petrochemical regions using the application of ANN- SOM. They collected data from 103 sampling sites of surface and underground water, sediments and soils distributed over three regions in the Niger Delta (Nigeria). Water (n = 11), soil (n = 38) and sediment (n = 54) samples were classified based on their chemical, toxicological and physical variables applying the SOM. The datasets analyzed in this study include physico-chemical variables , the sum of 7 carcinogenic polycyclic aromatic hydrocarbons (PAHs), the sum of 10 non-carcinogenic PAHs, sum of total petroleum hydrocarbon, the sum of benzene, toluene, ethyl benzene and xylene (BTEX), and 2 toxicity parameters (pH and EC).

The Mathworks software (Matlab) and SOM Toolbox version 2 was used for the SOM analysis (Vesanto et al, 1999). They used the SOM tool to project the input data (with multi-dimensions) into 2-dimensional lattice structure by going through a training phase and also preserving the topological features in the input data space. The computation began by running several iterations where the individual distance between the input vector and each weight vector (Euclidean distance) is calculated and the neuron weight with the shortest distance representative of an input vector is chosen as the best neuron called the best matching unit (BMU). The weights are updated at each time step, then the resultant weight vectors becomes the weighted averages of the vectors presented in the input space. In the training phase, the SOM exhibits adaptive flexibility by folding into the input data clouds (Vesanto et al., 1999). The performance which eventually generates similar neurons (BMUs) specifically located in the output space (on a grid in form of maps) based on the measured distances relative to the input vectors. From the result of the SOM dataset processing, it was observed that majority of the sites were contaminated with carcinogenic PAHs and carcinogenic heavy metals which are of concern due to the effects on human health. Comprehensive remediation and mitigation plans are recommended for these areas. Furthermore, the provision of effective health care facilities that can evaluate the health conditions of the residents in the high contamination zones identified by the SOM and the .provision of urgent care needed to those severely affected by the chronic exposure to these pollutants are also recommended.

Young-Seuk Park et al., (2014) investigated the water quality of 302 reservoirs distributed nationwide in Korea by classifying the using a self-organizing map (SOM), examining how hydrogeomorphometry variables are related to reservoir water quality, and evaluating the effects of variables at different categories including geology, land cover, hydro morphology, and physico-chemistry on reservoir water quality through a theoretical path model. The SOM classified the reservoirs into six clusters, from least to most polluted, with differences in physicochemical and hydrogeomorphometry variables between clusters. Water quality exhibits strong relationships with the proportions of urban, agricultural, and forest land cover types in the watersheds. Finally,they concluded that the results showed that hydrogeomorphometry of reservoirs and percentages of land cover types within watersheds have a considerable impact on the water quality of adjacent aquatic ecosystems.

An et al., (2016) used principal component analysis (PCA) and a self-organising map (SOM) to analyse a complex dataset obtained from the river water monitoring stations in the Tolo Harbor and Channel Water Control Zone (Hong Kong), covering the period of 2009–2011.They initially applied PCA to identify the principal components (PCs) among the nonlinear and complex surface water quality parameters. SOM followed PCA, and was implemented to analyze the complex relationships and behaviors of the parameters. The results reveal that PCA reduced the multidimensional parameters to four significant PCs which are combinations of the original ones. The positive and inverse relationships of the parameters were shown explicitly by pattern analysis in the component planes. It was found that PCA and SOM are efficient tools to capture and analyze the behavior of multivariable, complex, and nonlinear related surface water quality data.

IV. CONCLUSION AND DISCUSSION

Over the last decades, SOMs have increasingly been used for analysis, estimation and prediction of various problems in environmental engineering. These studies indicate that in many cases, SOM can exceed other methods. However, like feed-forward MLP applications, SOM applications are generally dependent on quick approaches characterized by guesswork and the trial - and- error approaches. Another interesting ability of SOMs is that they can be used to automatically group and/or typify data according to different properties. Likewise,

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

there are no proven techniques to assess the reliability or validity of SOM or feed forward MLP models. Thus, there are areas that need further consideration. There is a need to further investigate the ability of SOMs to describe and to further analyze non-linear processes in this field and to test the robustness of these methods in relation to traditional linear and non-linear approaches.

REFERENCES

1. Aguado D., Montoya, T., Borras, L., Seco, A., Ferrer, J.Using Self Organizing Maps and PCA analysis and interpreting data from a P-removal. Engineering apllication of Artificial intelligence. 2008.21.919-930.
2. Alhoniemi, E., Hollmen, J., Simula, O., Vesanto, J.,. Process monitoring and modeling using the self-organizing map. Integrated Computer-Aided Engineering. 1999. 6 (1), 3-14.
3. ASCE Task Committee on application of Artificial Neural Networks in Hydrology,. Artificial neural networks in hydrology. I: preliminary concepts. Journal of Hydrologic Engineering. 2000a . 5 (2), 115-123.
4. ASCE Task Committee on application of Artificial Neural Networks in Hydrology,. Artificial neural networks in hydrology. II: hydrologic applications. Journal of Hydrologic Engineering. 2000b. 5 (2), 124-137.
5. Berthold T., Peter Milbradt.Artificial Neuronal Networks in Environmental Engineering: Theory and Applications. 2009.18th International Conference on the Application of Computer Science and Mathematics in Architecture and Civil Engineering.
6. Bowden, G.J., Dandy, G.C., Maier, H.R.,. Input determination for neural network models in water resources applications. Part 1-background and methodology. Journal of Hydrology. 2005a .301, 75-92.
7. Bowden, G.J., Maier, H.R., Dandy, G.C.,. Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river. Journal of Hydrology. 2005b. 301, 93-107.
8. Dawson, C.W., Wilby, R.L.,. Hydrological modelling using artificial neural networks. Progress in Physical Geography. 2001. 25 (1), 80-108.
9. Heikkinen, M., Hiltunen, T., Liukkonen, M., Kettunen, A., Kuivalainen, R., Hiltunen, Y.,. A modelling and optimization system for fluidized bed power plants. Expert Systems with Applications. 2009. 36, 10274-10279.
10. Heikkinen, M., Kettunen, A., Niemitalo, E., Kuivalainen, R., Hiltunen, Y.,. SOM-based method for process state monitoring and optimization in fluidized bed energy plant. In: Duch, W., Kacprzyk, J., Oja, E., Zadro_zny, S. (Eds.), Lecture Notes in Computer Science. 3696. Springer-Verlag Berlin, Heidelber. 2005. 409 -414.
11. Heikkinen, M., Poutiainen, H., Liukkonen, M., Heikkinen, T. & Hiltunen, Y.Subtraction analysis based on self-organizing maps for an industrial wastewater treatment process. Mathametica. Computation Simulation. 2011. 82. 450–459.
12. Hilario Lopez Garcia, Ivan Machon Gonzalez. Self-organizing map and clustering for wastewater treatment monitoring. Engineering Applications of Artificial Intelligence. 2004. 17. 215–225.
13. Kohonen, T., 2001. Self-Organizing Maps. Springer-Verlag, Berlin.
14. Liukkonen, M., Heikkinen, M., Hälikkä, E., Kuivalainen, R., Hiltunen, Y.,. Emission analysis of a fluidized bed boiler by using self-organizing maps. In:Kolehmainen, M., Toivanen, P., Beliczynski, B. (Eds.), Lecture Notes in Computer Science 5495. Springer-Verlag, Berlin Heidelberg, 2009c. 119-129.
15. Liukkonen, M., Heikkinen, M., Hiltunen, T., Hälikkä, E., Kuivalainen, R., Hiltunen, Y., 2009b. Modeling of process states by using artificial neural networks in a fluidized bed energy plant. In: Troch, I., Breitenecker, F. (Eds.), Proceedings of MATHMOD 09 VIENNA. Argesim Publishing House, Vienna,pp. 2008.397-402.
16. M. Pavan, R. Todeschini. 2008. Scientific Data Ranking Methods: Theory and Applications, Elsevier, Amsterdam.
17. M.Liukkonen, T. Hiltunen, E.Halikka, Y. Hiltunen. Modeling of the fluidized bed combustion process of NO_x emission using SOM. An application to the diagnosis of process state. Environmenatl modeling and Software. 2011. 26, 605-614.
18. Maier, H.R., Dandy, G.C.,. The use of artificial neural networks for the prediction of water quality parameters. Water Resources Research. 1996. 32 (4), 1013-1022.
19. Maier, H.R., Dandy, G.C.,. Determining inputs for neural network models of multivariate time series. Microcomputers in Civil Engineering. 1997. 12 (5), 353-368.
20. Maier, H.R., Dandy, G.C.,. Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. Environmental Modelling and Software. 2000. 15, 101-124.
21. Richard Olawoyin,Antonia Nieto, Robert Larry Grason, Frank Hardisty, Samuel Oyewole,. Application of Artificial Neural Networks and Self organizing Maps for the categorization of water, soil and sediment quality in Petrochemical regions.Expert system with applications. 2013. 40,3634-3648.
22. Tsakovski, S., Simeonov, V., .Hasse diagram technique as exploratory tool in sediment pollution assessment, Journal of Chemometrics . 2011.25 (5) 254–261.
23. Tsvetomil Voyslavov, Stefan Tsakovski, Vasil Simeonov ,.Surface water quality analysis using Self organising maps and Hasse diagram technique. Chemometrics and Intelligent laboratory systems. 2012. 118, 280-286.
24. Vesanto, J., Alhoniemi, E., 2000. Clustering of the self-organizing map. IEEE Transactions on Neural Networks. 2000. 11 (3), 586-600.
25. Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J.,. Self-organizing map in Matlab: the SOM Toolbox. 1999. In Proceedings of the Matlab DSP conference, Espoo, Finland, Comsoloy.
26. Yan An, Zhihong Zou, Ranran Li, Descriptive.Characteristics of Surface Water Quality in Hong Kong by a Self-Organising Map. International Journal of Environmental Research and Public Health. 2016.13(1).115.
27. Young-Seuk Park , Yong-Su Kwon, Soon-Jin Hwang , Sangkyu Park.Characterizing effects of landscape and morphometric factors on water quality of reservoirs using a self-organizing map. Environmental modelling and software. 2014. 55. 214-221.