Application of the Self-organizing Map (SOM) to Characterize Nutrient Urban Runoff

Angela Gorgoglione[1(✉)], Alberto Castro[1], Andrea Gioia[2], and Vito Iacobellis[2]

[1] Universidad de la República, 11300 Montevideo, Uruguay
agorgoglione@fing.edu.uy
[2] Politecnico di Bari, 70126 Bari, Italy

Abstract. Urban stormwater runoff is considered worldwide as one of the most critical diffuse pollutions since it transports contaminants that threaten the quality of receiving water bodies and represent a harm to the aquatic ecosystem. Therefore, a thorough analysis of nutrient build-up and wash-off from impervious surfaces is crucial for effective stormwater-treatment design. In this study, the self-organizing map (SOM) method was used to simplify a complex dataset that contains precipitation, flow rate, and water-quality data, and identify possible patterns among these variables that help to explain the main features that impact the processes of nutrient build-up and wash-off from urban areas. Antecedent dry weather, among the rainfall-related characteristics, and sediment transport resulted in being the most significant factors in nutrient urban runoff simulations. The outcomes of this work will contribute to facilitating informed decision making in the design of management strategies to reduce pollution impacts on receiving waters and, consequently, protect the surrounding ecological environment.

Keywords: SOM · Nutrient urban runoff · Water quality · SWMM · Diffuse pollution

1 Introduction

Urban stormwater runoff is one of the principal causes that contribute to the non-point source (NPS) (or diffuse) pollution and subsequent impairments of rivers, streams, lakes, and estuaries [1, 2]. Diffuse nutrients are challenging to manage and reduce since their quantification is difficult to evaluate. In fact, it is not possible to identify a point source, and they are generated by the contribution of many small sources [3]. Surface-water eutrophication could produce the growth of phytoplankton biomass (causing the decrease in water transparency) and toxic and irritating algal blooms (with odor and taste issues). Consequently, an increment of the costs of drinking water treatment could be registered [4]. Hence, the sustainable management of urban watersheds plays a crucial role in the protection of the quality of surface water bodies.

The dynamic nature of urban stormwater quality has been attributed to several factors. Previous researches have demonstrated that pollutant concentrations observed in stormwater were highly sensitive to local environmental conditions. The latter can be

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12252, pp. 680–692, 2020.
https://doi.org/10.1007/978-3-030-58811-3_49

summarized as follows: *i)* temporal factors (like first flush pattern, antecedent dry period, rainfall trend) [5]; *ii)* spatial factors (like land use/land cover, watershed characteristics) [6]; *iii)* water-quality features (like pH, salinity, temperature, suspended solids) [7]; and *iv)* type of pollutants (like mass, composition, decay rates) [8]. In addition to the inherent randomness of nature, most of the works mentioned above have also shown that many multifaceted interactions among these factors have occurred in numerous dimensions [9].

During the last decade, linear multivariate statistical methods, like cluster analysis (CA) and principal component analysis/factor analysis (PCA/FA), have been well-reviewed in the evaluation and analysis of surface-water quality since their outstanding capability to process, analyze, and simplify a large amount of environmental data [3, 10, 11]. However, the linear multivariate statistical techniques are limited by the assumption of linearity, a hypothesis not verified for the process under study (pollutant build-up/wash-off).

Nowadays, the self-organizing map (SOM), a particular type of artificial neural network (ANN), has received attention from environmental researchers. SOM was proposed by Kohonen [12]. It is a competitive and unsupervised self-organizing network, formed of fully connected neuron arrays, which can produce the mapping from a multi-dimensional space to a two-dimensional space. With the pros of non-linear features, learning and induction skills, and a large number of parallel distributed structures, SOM has been successfully applied to tackle different environmental problems, such as water-quality characterization, water-area delineation with satellite images, and several other topics related to environmental engineering and water resources [13, 14].

Based on these considerations, this study aims to assess possible patterns among rainfall-runoff and water-quality variables that can help to analyze the main factors that influence nutrient build-up and wash-off from urban areas. We will take into account the non-linearity of these processes and the multidimensionality of the system involved. Therefore, to accomplish this objective, the SOM technique will be adopted.

The outcomes of this work not only will deliver useful information on emerging strategies for NPS management and regulation but also will widen the potential applications of SOM methodology in the environmental field, particularly in the water-resource engineering.

2 Materials and Methods

2.1 Methodology Description

The workflow reported in Fig. 1 summarizes the methodology adopted to accomplish the main objective of this work.

Three main phases can be identified: data set creation, data analysis, and results. More details about each of these phases can be found in the remainder of this paper. In particular, in Sect. 2.2, the data set creation is explained; in Sect. 2.4, all the techniques adopted are thoroughly described; in Sect. 3 and all its sub-sections, the results are rigorously presented and discussed.

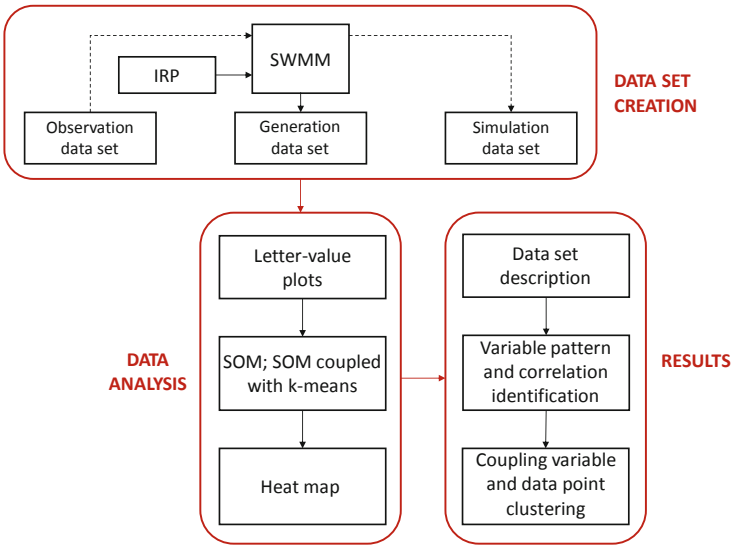


Fig. 1. Workflow summarizing the approach adopted in this study.

2.2 Study Area and Data Collection

The investigated area is represented by an urban watershed located in Sannicandro di Bari (hereafter called SB) (Puglia Region, Southern Italy). The catchment area is equal to 31.24 ha, including 21.87 ha (70%) of impervious surface. The SIT Puglia [15] land-use map indicates that the whole basin is residential, and only 3.80% of the watershed (1.20 ha) is characterized by a green area. The average slope of the catchment is equal to 1.56%, and the stormwater-drainage network is 1.96 km-long, collecting water into a concrete channel 1.20 m × 1.70 m [16]. The precipitation is recorded with a rain gauge located near the outlet of the basin. The water-quality monitoring is carried out through an autosampler with 24 bottles of 0.5 L each [16]. In Fig. 2, the watershed area, the drainage network, and the outfall are displayed.

In this work, three different datasets were used: *i) observations*: observed precipitation, flow rate, and water-quality data, used for calibrating and validating the hydrologic/hydraulic and water-quality models; *ii) generations*: flow rate and water-quality variables are obtained using generated precipitation time series, produced through numerical simulation, used as input in the hydrologic/hydraulic and water-quality models; *iii) simulations*: flow rate and water-quality data are obtained using observed precipitation as input data set in the hydrologic/hydraulic and water-quality models.

Considering the observations dataset, the monitoring campaign provided rainfall, flow rate, total suspended solids (TSS), and nutrients (N_{tot} and P_{tot}) for five rainfall events that occurred at SB (11/10/2006, 11/22/2006, 12/17/2006, 1/24/2007, and 2/10/2007). In Table 1, the observed dataset is summarized.

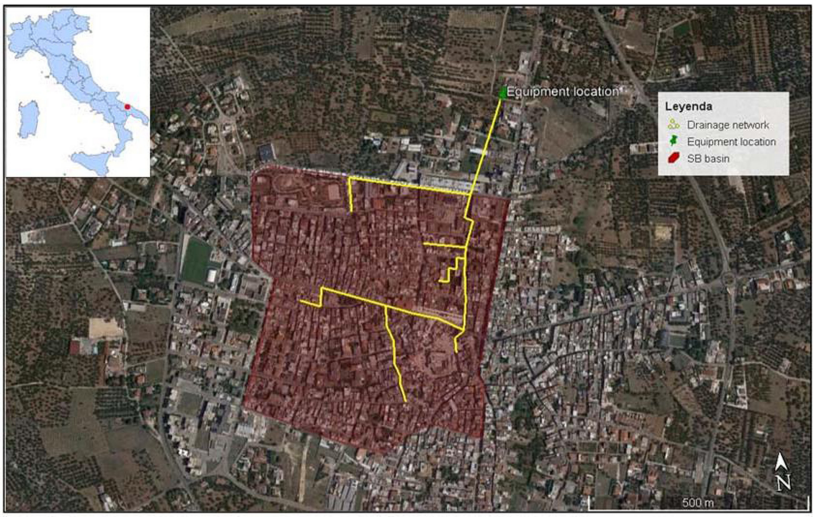


Fig. 2. Urban study area. Watershed surface (red area), drainage network (yellow lines), and location of the outfall (green pin). (Color figure online)

Table 1. Summary of the (a) rainfall-runoff and (b) water-quality data for the monitored events at SB watershed.

(a)	Event	ADP (days)	Total rainfall (mm)	Event duration (min)	Max. Rainfall intensity (mm/h)	Runoff volume (m^3)	Runoff peak (m^3/s)
	11/10/2006	6	2.40	50.00	24.00	113.49	0.04
	11/22/2006	11	4.30	112.00	6.00	148.86	0.04
	12/17/2006	18	5.90	251.00	12.00	286.88	0.05
	1/24/2007	19	1.60	37.00	12.00	111.62	0.05
	2/10/2007	6	12.90	398.00	36.00	460.11	0.05

(b)	Event	TSS (mg/L)			N_{tot} (mg/L)			P_{tot} (mg/L)		
		min	max	EMC	min	max	EMC	min	max	EMC
	11/10/2006	224.0	420.0	19.54	7.0	8.3	0.47	0.70	1.00	0.05
	11/22/2006	124.0	2160.0	86.40	3.6	14.0	0.45	0.24	2.96	0.11
	12/17/2006	6.0	217.0	6.040	NA			NA		
	1/24/2007	177.0	807.0	47.96	5.4	10.0	0.48	0.65	0.99	0.03
	2/10/2007	541.0	2090.0	40.00	6.3	13.0	0.25	2.08	3.63	0.08

Regarding the *generation* dataset, synthetic precipitation time series were produced exploiting the Iterated Random Pulse (IRP) rainfall model (proposed by Veneziano and Iacobellis [17]), providing a time series of 15 years length and 15 min of aggregation. More details regarding the model description, its parameters, and its implementation to the study area is provided in the literature [18, 19]. Taking into account the regional regulation [20], single rainfall events were defined considering a condition of 48 h of antecedent dry weather. Accordingly, 567 rainfall events were identified in the 15-years-time series.

Regarding the *simulation* dataset, observed rainfall events and simulation results of runoff and water quality were considered. A brief description of the adopted rain-fall-runoff and water quality models is reported in Sect. 2.2. The model calibration and validation was an issue already tackled in our previous work [16].

2.3 Hydrologic/Hydraulic and Water-Quality Model Implementation

The Environmental Protection Agency's Storm Water Management Model (EPA's SWMM) was adopted in this work to estimate the flow rate and water quality variables [21]. In particular, the runoff block was used to simulate water runoff, the formation of surface runoff constituent loads due to the pollutant build-up during dry weather, and pollutant wash-off during wet weather. The transport block was also used to execute the flow and pollutant routing through the drainage network. In the transport module, flow routing was accounted for utilizing the kinematic-wave; in turn, water-quality dynamics included first-order degeneration within the sewer system [22]. Furthermore, the accounted water losses are represented by the depression storage on the impervious surface of the watershed and the infiltration amount. In this work, Horton's equation was used to calculate the infiltration rate, whose parameters have been estimated considering the typical values reported in the recent literature [23].

A comprehensive explanation of the mathematical description of these physical processes is reported in the recent bibliography [16, 22]. In particular, Di Modugno et al. [16] described good performances regarding the evaluation of the water quantity response of SB by exploiting a hydrologic/hydraulic approach with several sets of input parameters; moreover, the build-up and wash-off simulations were positively performed.

2.4 Data Analysis

The SOM is a particular kind of ANN that learns in an unsupervised manner, as there is no target or objective to compare with [24]. It is conceived to self-manage similar information that has not yet been classified. In the Self-Organizing Map, neurons compete with each other in order to describe the input data, as opposed to error-correction learning (such as backpropagation with gradient descent). As an outcome, data in the multidimensional attribute space can be reorganized to a smaller number of latent dimensions, which is arranged considering a predetermined geometry in a space of lower dimensionality, usually an ordinary two-dimensional array of neurons.

K-means CA was exploited to organize the investigated variables into clusters base on their similarity. A distance function evaluates the similarity, firstly, among data points, afterward, among the groups as well [25]. This approach also needs a priori selection of an arbitrary number of groups (k). In this work, the silhouette method was employed to evaluate the most suitable k [26]. This approach measures the proximity of each point of a cluster to the points of neighboring clusters:

$$s(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} \tag{1}$$

where *a(i)* is the average distance of the point *(i)* with respect to all the other points in the assigned cluster *(A)*, *b(i)* is the average distance of the point *(i)* with respect to other points belonging to its closest neighboring cluster *(B)*. Silhouette values are within the range [−1, 1]; the higher the value, the better is the cluster conformation.

Heat map analysis is a pseudo-color picture, including two dendrograms for two different objects, which can be divided into several clusters [27]. The different influence characteristics in these two objects were reorganized considering their similarity based on CA.

3 Results and Discussion

3.1 Data Set Description

Figure 3 shows the distribution and quantiles of the investigated water-quality variables, obtained using the three different given datasets (generations, simulations, and observations). In particular, each variable was normalized based on its mean and standard deviation, so that the new standardized variable is characterized by mean equal to 0 and standard deviation equal to 1. This plot style is known as a "letter-value plot" because it shows a large number of quantiles. It is similar to a box-plot in plotting a non-parametric representation of a distribution in which all features correspond to actual observations [28]. In particular, the horizontal gray lines denote the median, and the colored dots correspond to extreme values.

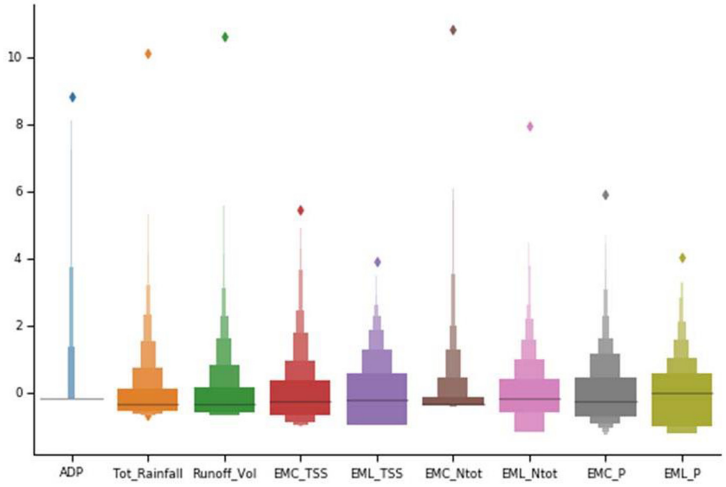


Fig. 3. Letter-value plots of the complete dataset (observations, simulations, and generations). The horizontal gray line represents the median; the colored dot corresponds to outliers.

In Table 2, the main quantiles (25, 50, and 75%) and the maximum values of the standardized variables are reported.

The letter-value plots reported in Fig. 3 are able to show the skewed tails of the investigated variables. In particular, the figure highlights the strong asymmetry of the standardized variables, except for the *EML_P* and *EML_TSS*, which show lower asymmetry.

Table 2. Statistical description of rainfall-runoff and water-quality variables.

	ADP	Tot_Rainfall	Runoff_Vol	EMC_TSS	EML_TSS	EMC_N_{tot}	EML_N_{tot}	EMC_P	EML_P
mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
sd	1.001	1.001	1.001	1.001	1.001	1.000	1.000	1.000	1.000
min	−0.177	−0.680	−0.674	−0.993	−0.968	−0.414	−1.182	−1.247	−1.227
25%	−0.176	−0.565	−0.575	−0.680	−0.959	−0.378	−0.593	−0.724	−1.005
50%	−0.175	−0.329	−0.327	−0.267	−0.198	−0.334	−0.196	−0.256	0.000
75%	−0.173	0.133	0.139	0.349	0.577	−0.138	0.407	0.438	0.565
max	8.841	10.136	10.608	5.442	3.919	10.816	7.964	5.893	4.025

3.2 Variable Correlations

The visualization of the weight maps is a useful tool to identify possible correlations of the different rainfall-runoff and water-quality variables (Fig. 4). To build the map, initially, the input data is normalized per variable. Then, the map size is evaluated by calculating the number of neurons from the number of data points in the training data using the following equation [29]:

$$M \approx 5\sqrt{N} \tag{2}$$

where M is the number of neurons, which is an integer close to the result of the right-hand side of the equation, and N is the number of data points. In this study, there are 577 input data points (N), so the number of neurons used was 121 (M*),* i.e., a map of 11 by 11 neurons. In this work, the implementation was coded in Python on a 2.6 GHz Intel i7 PC with 32 GB of memory using the *minisom* package for the creation of the SOM [30], and the *seaborn* package for visualization [31]. The phases of the map-weights initialization and the map training were both realized picking samples at random from the input data set. After a training phase of 10,000 epochs (all the input data samples were used 10,000 times), a quantization error of 0.095 and a topographic error of 0.067 were obtained, assuring the quality of the resulting map.

Considering the rainfall-related variables (*ADP*, *Tot_Rainfall*, and *Runoff_Vol.*), the three weight maps presented in Fig. 4 show consistent results. *Tot_Rainfall* and *Runoff_Vol.* activate the same neurons (positive neurons), which are symmetric to the neurons activated by *ADP*. The "redder" the activated neurons of *ADP* are, the greater the water loss in the system, the "bluer" are the negative neurons of runoff volume and total rainfall maps.

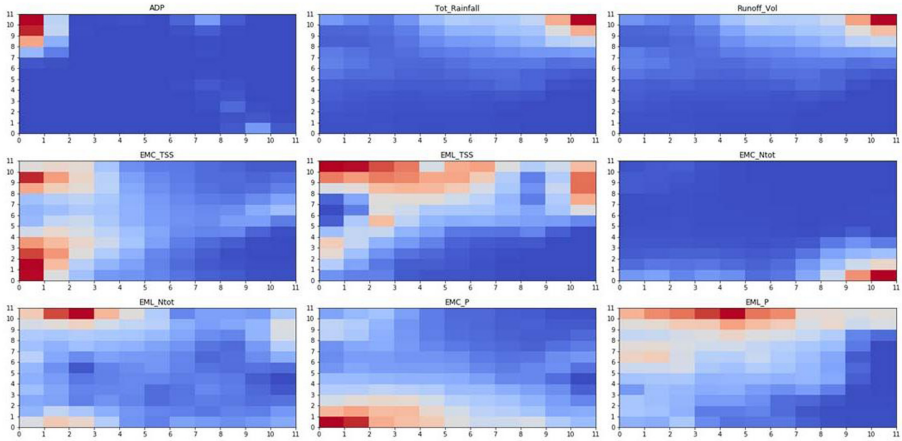


Fig. 4. Pattern evaluation of the rainfall and water-quality variables with the SOM-weight map.

Among the water-quality-related features, three map patterns can be identified: pollutant EMLs, *EMC_TSS* and *EMC_P*, and *EMC_*N_{tot}. One of the significant findings achieved with this analysis is the strong relationship between sediment load (*EML_TSS)* and nutrient EMLs and between the sediment and phosphorus concentration (*EMC_TSS* and *EMC_P)*. This correlation suggests that sediment transport plays an essential role in the mobilization of nutrients from urban impervious surfaces. In particular, in our study area, phosphorus transport is highly correlated to the sediment wash-off. The particle-bound portion of the nitrogen is depicted by the same positive activated neurons of *EML_TSS* and *EML_*N_{tot}. While its dissolved portion is represented by the symmetric positive neurons between *EMC_*N_{tot} and *Tot_Rainfall-Runoff_Vol*. This correlation suggests that the bigger the runoff volume, the smaller the nitrogen concentration. The latter can be justified with the dilution process. Therefore, in our study area, nitrogen mobilization predominantly occurs in the dissolved form. Furthermore, it is worth remarking that the red positive neurons activated by *ADP* are located in the high-left corner, as well as those activated by pollutant EMLs. Therefore, the longer the antecedent dry weather, the higher is the amount of pollutants built up on the impervious surface, mainly sediments.

It is worth noting that the water quality-related features activate more neurons than the rainfall-related ones, that are, instead, located in a single area and not sparse all over the map. This means that the variation of the water-quality variables is higher and the neighbor region is ready to acquire their information.

3.3 SOM/Clustering Results

The analysis was performed by a SOM and k-means algorithm ensemble to group the variables into clusters based on their similarities. To identify the most representative number of clusters (k), the silhouette method was adopted. The average silhouette score of all the samples in the data set was calculated (Fig. 5a). This produces a value that

represents the silhouette score of that particular cluster. Furthermore, the boxplot of silhouette scores was built to assess the dispersion of the data at each number of clusters taken into account. (Fig. 5b). We considered the possibility of having from 2 to 10 clusters ($k \in [2, 10]$).

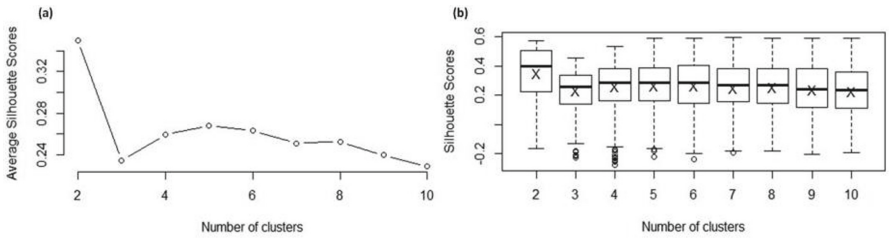


Fig. 5. (a) Average silhouette scores per k; (b) boxplot of silhouette scores per k (with $k \in [2, 10]$). The average is represented by 'X'.

In Fig. 5a, the optimum number of clusters seems to be $k = 2$. However, in Fig. 5b, we can see that the data dispersion related to $k = 2$ is higher than the rest. Moreover, the $k = 3$ and $k = 4$ have several outliers under the minimum value. Therefore, we decided to group the data points into 5 clusters, also considering that this boxplot has the highest maximum, and the average silhouette scores are the second-highest after $k = 2$. In Fig. 6(a), a matrix that shows the neighboring distances is represented (SOM-distance map). The higher the difference between one node and its neighbor, the darker the resulting color between those two neurons. The SOM outcome was coupled with k-means cluster analysis. In Fig. 6(a), the five clusters of the neurons are overlapped on top of the distance map and represented with different shapes and colors (green cross, orange x, blue circle, purple rhombus, and red square). Considering that each neuron is a vector, in Fig. 6(b), the feature with the highest value in each neuron is highlighted.

From Figs. 6(a) and (b), it is possible to see that the "green cross" group is characterized by high values of EMC_N_{tot}. The "orange x" cluster has mainly high values of *EMC_P* and also EMC_N_{tot} and *EMC_TSS*. The other three groups are more heterogeneous. The "red square" one is characterized by high values of *EMC_TSS, EML_P, ADP, EML_Ntot, EML_TSS*. The "purple rhombus" cluster has high values of *Tot_Rainfall*, *Runoff_Vol.*, *EML_TSS*, *ADP*, EMC_N_{tot}, *EMC_TSS*. The "blue circle" group, the biggest, is characterized by high values of all the variables.

With the aim of coupling variable-clustering and data point-clustering, a heat map analysis was run using Ward linkage and Euclidean distance (Fig. 7).

Also for the heat map, we considered 5 clusters ($k = 5$). The left dendrogram was obtained considering the original matrix (577×9), being 577 generated, observed, and simulated events and 9 rainfall and water-quality variables. It is possible to highlight that *ADP* is playing an essential role by clearly marking the 5 clusters. Furthermore, we can appreciate that for groups with high values of *Tot_Rainfall* and *Runoff_Vol*, the values of EMCs are low and vice-versa. The latter is explained by the dilution process.

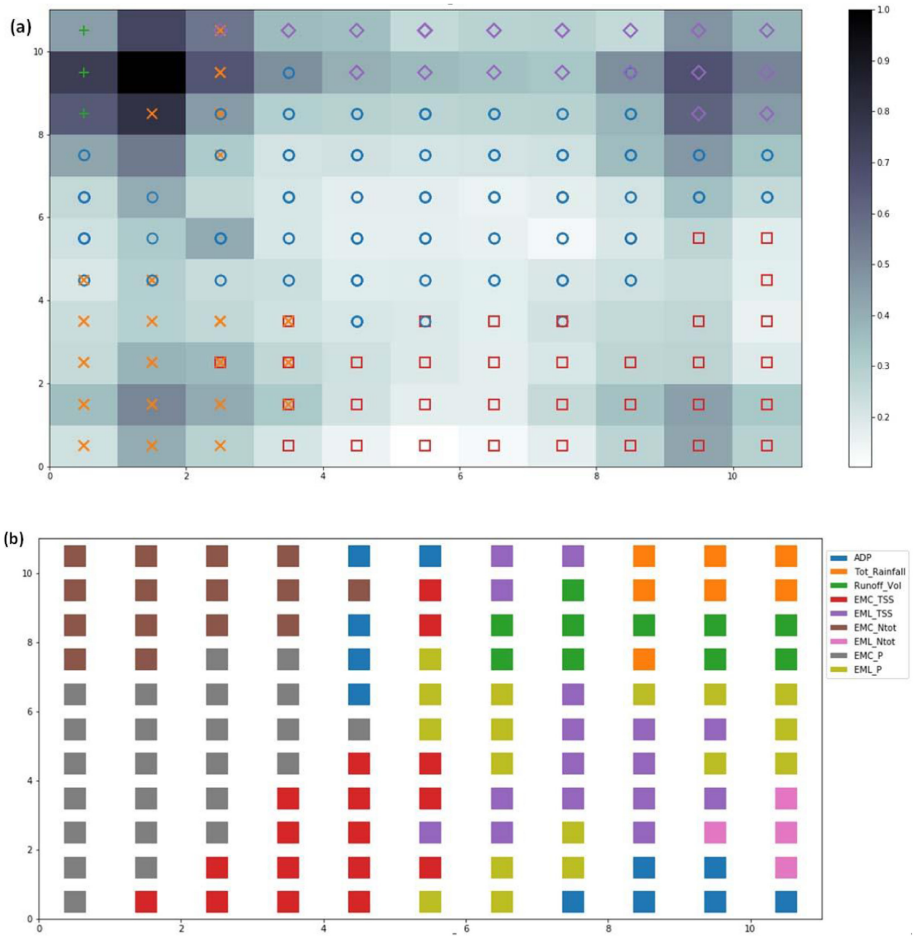


Fig. 6. (a) SOM-distance map coupled with k-means. (b) Variables with the highest value in each neuron. (Color figure online)

The top dendrogram was obtained considering the transposed of the previous matrix (9×577). The first cluster identified is formed by *Tot_Rainfall* and *Runoff_Vol*. *ADP* has its own cluster. EMLs are all grouped together. *EMC_TSS* and *EMC_P* are grouped, while *EMC_*N_{tot} forms another cluster. This portion of the heat map perfectly confirms the outcomes obtained with the SOM-weight map (Fig. 4).

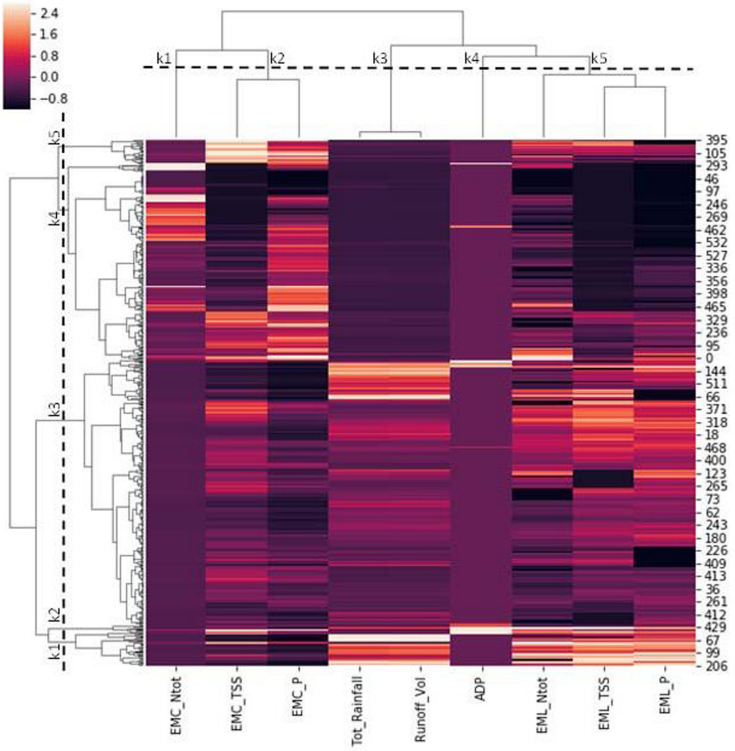


Fig. 7. Heat map analysis outcome (Ward linkage and Euclidean distance).

4 Conclusions

In this study, patterns among rainfall-runoff and water-quality variables were evaluated to identify the main factors that influence nutrient build-up and wash-off from urban areas, taking into account the non-linearity of these processes and the multidimensionality of the system involved. A study area located in southern Italy and the SOM technique were considered for this study. The main conclusions can be summarized as follows:

- Rainfall factors: *ADP*, *Tot_Rainfall*, and *Runoff_Vol.* present a reliable correlation driven by *ADP*. In fact, the "redder" the activated neurons of *ADP* are, the higher the water loss in the system, the "bluer" are the negative neurons of runoff volume and total rainfall maps.
- Water-quality factors: *EML_TSS* and nutrient EMLs present a robust correlation, confirming as appropriate the use of suspended solids as a proxy for the study of the behavior of nutrients in urban areas. Considering pollutant EMCs, a strong relationship was found only between the *EMC_TSS* and *EMC_P,* showing the high importance of the phosphorus particle-bound in the system under study.

- Interactions rainfall and water-quality factors: the nitrogen dilution process is represented by the symmetric positive neurons between *EMC_*N_{tot} and *Tot_Rainfall-Runoff_Vol*. A high correlation was also found between *ADP* and pollutant EMLs: the longer the antecedent dry weather, the higher is the amount of pollutants built up on the impervious surface, mainly sediments.

The outcomes of this work have proved that coupling watershed-scale studies and non-linear exploratory analysis, like SOM, can provide an adequate overview of the relationships between water-quality variables and rainfall characteristics. Furthermore, the results presented in this work are expected to assist researchers and technicians in quantifying their confidence in the water-quality assessment, which aids informed analysis and decision-making in the design of management strategies to reduce pollution impacts on receiving waters and, consequently, protect the surrounding ecological environment.

References

1. Bernhardt, E.S., Palmer, M.A.: Restoring streams in an urbanizing world. Freshw. Biol. **52**, 738–751 (2007)
2. Jiang, S.C., Lim, K.-Y., Huang, X., McCarthy, D., Hamilton, A.J.: Human and environmental health risks and benefits associated with use of urban stormwater. Wiley Interdiscip. Rev. Water **2**, 683–699 (2015)
3. Gorgoglione, A., Gioia, A., Iacobellis, V.: A framework for assessing modeling performance and effects of rainfall-catchment-drainage characteristics on nutrient urban runoff in poorly gauged watersheds. Sustainability **11**, 4933 (2019)
4. Brett, M.T., et al.: Non-point-source impacts on stream nutrient concentrations along a forest to urban gradient. Environ. Manage. **35**, 330–342 (2005)
5. Lee, H., Lau, S.L., Kayhanian, M., Stenstrom, M.K.: Seasonal first flush phenomenon of urban stormwater discharges. Water Res. **38**(19), 4153–4163 (2004)
6. Gobel, P., Dierkes, C., Coldewey, W.C.: Storm water runoff concentration matrix for urban areas. J. Contam. Hydrol. **91**(1–2), 26–42 (2007)
7. Nguyen, H.L., Leermakers, M., Elskens, M., De Ridder, F., Doan, T.H., Baeyens, W.: Correlations, partitioning and bioaccumulation of heavy metals between different compartments of Lake Balaton. Sci. Total Environ. **341**(1–3), 211–226 (2005)
8. Surbeck, C.Q., Jiang, S.C., Ahn, J.H., Grant, S.B.: Flow fingerprinting fecal pollution and suspended solids in stormwater runoff from an urban coastal watershed. Environ. Sci. Technol. **40**(14), 4435–4441 (2006)
9. Ki, S.J., et al.: Advancing assessment and design of stormwater monitoring programs using a self-organizing map: characterization of trace metal concentration profiles in stormwater runoff. Water Res. **45**, 4183–4197 (2011)
10. Varol, M.: Spatio-temporal changes in surface water quality and sediment phosphorus content of a large reservoir in Turkey. Environ. Pollut. **259**, 113860 (2020)
11. Batur, E., Maktav, D.: Assessment of surface water quality by using satellite images fusion based on PCA method in the Lake Gala, Turkey. IEEE Trans. Geosci. Remote Sens. **57**(5), 2983–2989 (2019)
12. Kohonen, T.: Automatic formation of topological maps of patterns in a self-organizing system. In: Oja, E., Simula, O. (eds.) Processing 2nd Scandinavian Conference on Image Analysis, pp. 214–220 (1981)

13. Sengorur, B., Koklu, R., Ates, A.: Water quality assessment using artificial intelligence techniques: SOM and ANN—a case study of Melen River Turkey. Water Qual. Exposure Health **7**(4), 469–490 (2015)
14. Jiang, M., Wang, Y., Tang, Q., Meng, F., Yao, Z., Cheng, P.: Assessment of surface water quality using a growing hierarchical self-organizing map: a case study of the Songhua River Basin, northeastern China, from 2011 to 2015. Environ. Monit. Assess. **190**, 260 (2018). Regional geographical information system: SIT Puglia: Available online: http://www.sit.puglia.it/. Accessed 8 Jan 2020
15. Di Modugno, M., et al.: Build-up/wash-off monitoring and assessment for sustainable management of first flush in an urban area. Sustainability **7**, 5050–5070 (2015)
16. Veneziano, D., Iacobellis, V.: Multiscaling pulse representation of temporal rainfall. Water Resour. Res. **38**, 131–1313 (2002)
17. Veneziano, D., Furcolo, P., Iacobellis, V.: Multifractality of iterated pulse processes with pulse amplitudes generated by a random cascade. Fractals **10**, 209–222 (2002)
18. Gorgoglione, A., Gioia, A., Iacobellis, V., Piccinni, A.F., Ranieri, E.: A rationale for pollutograph evaluation in ungauged areas, using daily rainfall patterns: Case studies of the Apulian region in Southern Italy. Appl. Environ. Soil Sci. 2016 (2016)
19. Regional Regulation: Stormwater runoff and first flush regulations (implementation of article 13 of Legislative Decree no 152/06 and subsequent amendments), no 26, 9 December 2013. https://www.indicenormativa.it/sites/default/files/R_26_09_12_2013.pdf. Accessed 6 Jan 2020
20. Rossman, L.A.: Storm water management model user's manual version 5.0; EPA/600/R-05/040; National Risk Management Research Laboratory-Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH, USA (2010)
21. Gorgoglione, A., Bombardelli, F.A., Pitton, B.J.L., Oki, L.R., Haver, D.L., Young, T.M.: Uncertainty in the parameterization of sediment build-up and wash-o processes in the simulation of sediment transport in urban areas. Environ. Model. Softw. **111**, 170–181 (2019)
22. Horton, R.E.: An approach toward a physical interpretation of infiltration capacity. Soil Sci. Soc. Am. **5**, 399–417 (1940)
23. Kohonen, T.: Self-organizing Maps, 3rd edn. Springer, Heidelberg (2001). https://doi.org/10.1007/978-3-642-56927-2
24. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surv. **31**, 264–323 (1999)
25. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
26. Wei, C., Gao, C., Han, D., Zhao, W., Lin, Q., Wang, G.: Spatial and temporal variations of water quality in Songhua River from 2006 to 2015: implication for regional ecological health and food safety. Sustainability **9**, 1502 (2017)
27. Hofmann, H., Kafadar, K., Wickham, H.: Letter-value plots: boxplots for large data. J. Comput. Graph. Stat. **26**(3), 469–477 (2017)
28. Vesanto, J.: SOM implementation in SOM toolbox. SOM toolbox online help. http://www.cis.hut.fi/projects/somtoolbox/documentation/somalg.shtml. Accessed Feb 2020
29. Vettigli, G.: Minisom: minimalistic and numpy-based implementation of the self organizing map. https://github.com/JustGlowing/minisom. Accessed Feb 2020
30. Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Ostblom, J., Lukauskas, S., et al.: mwaskom/seaborn: v0.9.0, July 2018. https://www.doi.org/10.5281/zenodo.1313201. Accessed Feb 2020