Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 89 (2014) 648 – 655

Procedia Engineering

www.elsevier.com/locate/procedia

16th Conference on Water Distribution System Analysis, WDSA 2014

Artificial Neural Networks and Entropy-Based Methods to Determine Pressure Distribution in Water Distribution Systems

E. Ridolfi[a,b,*], F. Servili[a], R. Magini[a], F. Napolitano[a], F. Russo[a], L. Alfonso[c]

[a]*Dipartimento di Ingegneria Civile, Edile e Ambientale, Sapienza University of Rome, 00184 Rome, Italy*
[b]*H2CU - Honors Center of Italian Universities, Sapienza University of Rome, 00184 Rome, Italy*
[c]*Department of Integrated Water Systems and Governance - Hydroinformatics, UNESCO-IHE Institute for Water Education, Delft, The Netherlands*

Abstract

Pressure determination in water distribution systems (WDS) is important because it generally drives the operational actions for leakage and failure management, backwater intrusion and demand control. This determination would ideally be done through pressure monitoring at every junction in the distribution system. However, due to limited resources, it is only possible to monitor at a limited number of nodes. To this end, this work explores the use of an Artificial Neural Network (ANN) to estimate pressure distributions in a WDS using the available data at the monitoring nodes as inputs. The optimal subset of monitoring nodes are chosen through an entropy-based method. Finally, pressure values are compared to synthetic pressure measures estimated through a hydraulic model.

Peer-review under responsibility of the Organizing Committee of WDSA 2014
Keywords: Water distribution system; water demands; scaling laws; Artificial Neural Network; entropy

1. Introduction

The issue of water pressure determination in a water distribution system (WDS) is significant when dealing with operational and design actions. Indeed, the knowledge of pressure values allows for proper management of the system, including leakage, water intrusion and demand control.

Unfortunately, pressure monitoring sensors are often placed at some nodes only, leaving a high uncertainty about pressure value in all the others. As limited resources do not allow for installing sensors in all nodes, many authors [e.g. 1] have dealt with the issue of determining the best sensors locations, by sampling design (SD). The problem can be solved by finding the best trade-off between calibrated model accuracy and the SD cost [e.g. 2,3].

The aim of this work is to propose a methodology to determine water pressure data series at unknown nodes, coupling an Artificial Neural Network with an entropy-based method.

* Corresponding author. Tel.: +39-064-458-5063 ; fax: +39-064-458-5063.
E-mail address: elena.ridolfi@uniroma1.it

Peer-review under responsibility of the Organizing Committee of WDSA 2014
doi:10.1016/j.proeng.2014.11.490

The entropy concept was first introduced in the thermodynamic field to describe the characteristic of all systems transformations to evolve toward the condition of highest disorder. Lately, it was defined by [4] as the measure of uncertainty about the knowledge of the state of a given system. The entropy concept formulated by [4] has been widely applied in water resources fields. Among all applications, it is used as a tool to optimize hydraulic, hydrological and water resources networks. For instance, using an entropy-based method, [5] evaluated the optimum number of raingauges, [6] designed a groundwater wells network, [7] determined the optimal density of raingauges over an area where an existing network was already deployed, [8–10] optimized water level monitoring sensors, [11] optimized the number of river cross sections in a 2D environment. In the water distribution systems environment, [12] presented three SD methods to select pressure monitoring locations for estimating pipe roughness coefficients. The first two approaches use the shortest path algorithm logic method, while the third is based on the entropy concept. Specifically, they maximize the entropy values using a genetic algorithm search method. However, there still remained the issue of calibration parameters values that need to be determined before the optimization-based SD procedure. [13] overcame the issue posing it as a two-objective optimization problem maximizing the calibrated model accuracy and minimizing the cost by reducing the number of sensors. [14] optimally locate water pressure gauge by determining the pressure change in each node when the demand is changing in one node using the entropy concept.

The main idea of this work is to determine from the most representative nodes of the network the pressure values in all others. To determine whether a node is representative or not the entropy concept is used. First, through the use of EPANET [15] pressure values are estimated using a Monte Carlo simulation [16] from water demand scenarios derived with the scaling laws approach [17–19].

Second, the most informative and least redundant nodes among all are found solving the Multi-Objectives Optimization Problem [MOOP; 10,11]. The two objective functions are the maximization of the information provided by the chosen group of nodes (represented by the Joint Entropy - JH) and the minimization of the their redundancy (measured by the Total Correlation - ToTC). The optimal sets of solutions are plotted on a 2D Pareto front. Then, the pressure values of the optimal groups (i.e. the optimal solutions of the MOOP) are used as input of the ANN. As output the pressure values at other nodes are computed. Finally, these generated pressure values are compared with those simulated through the hydraulic model EPANET. Indeed, in absence of a suitable data set, these simulated values are assumed to be truth (“observed”). Results show that the entropy-based methods, coupled with ANN approach, has a great potential to estimate water pressure from a limited number of observations. It means that it is possible to reduce the number of pressure monitoring sensors in a WDS, without reducing the monitoring quality.

2. Evaluation of the most representative pressure locations

To determine the marginal entropy of a discrete random vector (RV) X, it is necessary to define a partition of X. Let us consider that each component x_i of the RV has a probability $P(X = x_i) = p(x_i)$ of occurrence and the union of all x_i gives the certain event. Thus, they are a partition U_X of X and its entropy is [20]:

$$H(U_X) = H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i), \tag{1}$$

where n is the length of the RV. The concept can be generalized to define the joint entropy of N discrete random vectors RVs:

$$H(X_1, X_2, ..., X_N) = -\sum_{i_1}^{n_1} ... \sum_{i_N}^{n_N} p(i_1, ..., i_N)\log_2 p(i_1, ..., i_N), \tag{2}$$

where $p(i_1, ..., i_N)$ is the joint probability of the N variables. In our case X represent the location of a pressure sensor with an associated time series of n records. Therefore, the Multi-Objective Optimization Problem can be expressed as [9]:

$$\begin{aligned} max(JH) &= max\{H(X_1, X_2, ..., X_N)\} \\ min(ToTC) &= min\{ToTC(X_1, X_2, ..., X_N)\}, \end{aligned} \tag{3}$$

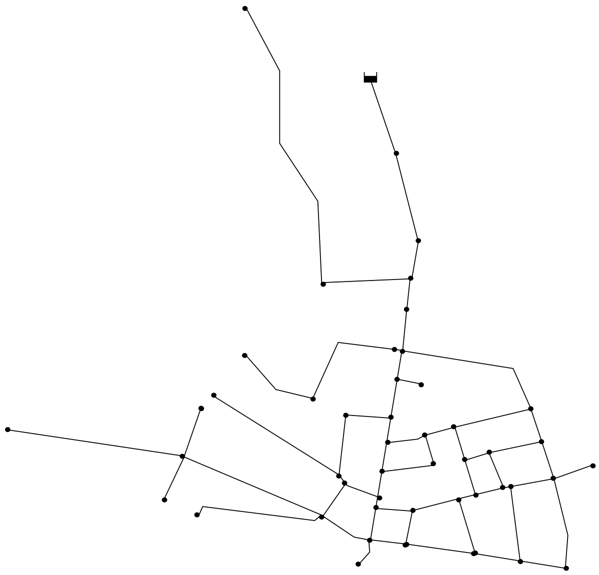

Fig. 1. Piedimonte San Germano water distribution network, Italy

where ToTC estimates the information shared between the N variables:

$$ToTC = \sum_{i=i}^{N} H_i - H(X_1, X_2, ..., X_N). \quad (4)$$

The solutions of Eq. 3 depict the most informative and least redundant nodes of the WDS. Therefore, each group of N nodes represent the optimal location where deploy the pressure monitoring sensors.

Terminal nodes are not used as input to the MOOP, because they are linked to one node only and cannot be considered representative.

3. Case study

The water distribution network considered in this work is located in Piedimonte San Germano (IT). It comprises 47 nodes, 60 pipes either in cast iron or in plastic and a reservoir, Figure 1. The elevation of all WDS nodes and the reservoir are reported in Table1.

Demand scenarios are generated characterizing the nodal demands over all the network. The water demand is modelled as a stochastic variable, and demands of different users are assumed to be correlated. The single-user values of the demand parameters were taken from real demand data series of Latina, Italy [21], because of the similarity of users' characteristics. The considered statistic values are: mean=0.48 l/min and variance=1.16 l/min, the cross-correlation coefficient equals 0.008.

3.1. Data analysis: demands scenarios and corresponding water pressures

Values of the mean, variance, cross-covariance and cross-correlation coefficients of the aggregated demand at each node are inferred resorting to scaling laws calibrated for the demand data of Latina [22]. From these data, correlated nodal demand series are generated using the Multivariate Streamflow Model by Fiering [23]. The scaling laws are used to estimate the values of demands statistics for any number of users, on the base of the demand signal of a single user. For the sake of brevity they will not be reported here, an extensive explanation can be found in [24]. Then, a Monte Carlo simulation is used to determine 1500 water demand scenarios.

Table 1. Piedimonte WDS nodes elevation.

Node ID	Elevation [m]	Node ID	Elevation [m]	Node ID	Elevation [m]	Node ID	Elevation [m]
1	138	**13**	115	**25**	107	**37**	109
2	129	**14**	109	**26**	107	**38**	109
3	126	**15**	114	**27**	110	**39**	109
4	128	**16**	118	**28**	111	**40**	108
5	157	**17**	119	**29**	112	**41**	108
6	125	**18**	112	**30**	114	**42**	108
7	120	**19**	108	**31**	114	**43**	110
8	120	**20**	107	**32**	115	**44**	110
9	118	**21**	107	**33**	118	**45**	111
10	116	**22**	106	**34**	117	**46**	111
11	111	**23**	105	**35**	111	**47**	122
12	119	**24**	104	**36**	110	**Reservoir**	280

Once that water demands scenarios are known at each node, corresponding pressure values are estimated using EPANET [15].

Since entropy estimation requires the definition of marginal and joint probability values, pressure values time series are analysed. Specifically, to estimate marginal probabilities, each RV (i.e. each node) has been divided in class intervals and the frequency of each class has been computed. Joint probabilities are evaluated using the grouping property of mutual information [25], extensively explained by [9].

4. The ANN based method

The Artificial Neural Network is a simple network formed by three layers. The first is also named input layer. Input data are represented by the pressures distribution in the distribution system at the known nodes. For every nodes we have 1500 scenarios of pressure. The second layer, named hidden layer, can be formed by several layers. In the considered ANN, for this layer we have chosen a number of 100 neurons. Finally, the third layer is named output layer, where the simulated outputs from the ANN are obtained. In this case, the outputs are the pressures in the remaining ungauged nodes of the network, evaluated from the pressures in the known nodes. The ANN needs a training phase, where the net internal processes, weights and the bias among the neurons of the ANN are calibrated to provide as output the best pressure values. This training process is performed using a percentage of pressure data series of all nodes. Specifically, 70% of pressure data set is used to train the ANN, 15% is used to validate its results, 15% is used to test them.

5. Results and discussions

From the entropy-based analysis the most representative nodes of the network are plotted on a 2D Pareto front, Figure 2, whose axes represent the two objective functions presented in Eq. 3. From the Figure it is possible to observe that even considering a small amount of nodes, the JH of optimal solutions is very close to its maximum value, while ToTC increases consistently as the number N increases as well. This means that a small amount of nodes might be used to represent the pressure values of the whole WDS. When more than 12 nodes are considered, the JH is constant, while ToTC continues increasing. The gain in terms of information using 12 nodes rather than a lower number is not as high as the loss of benefit for the high redundancy.

Once that the most representative group formed by N nodes are found (here N=3, 4, 5, 8, 10, 12, 15, 17), from their pressure values, the ANN is used to estimate the pressure distributions of the other nodes. To validate the methodology the errors between simulated and observed pressure values are estimated using the Mean Absolute Error (MAE) and

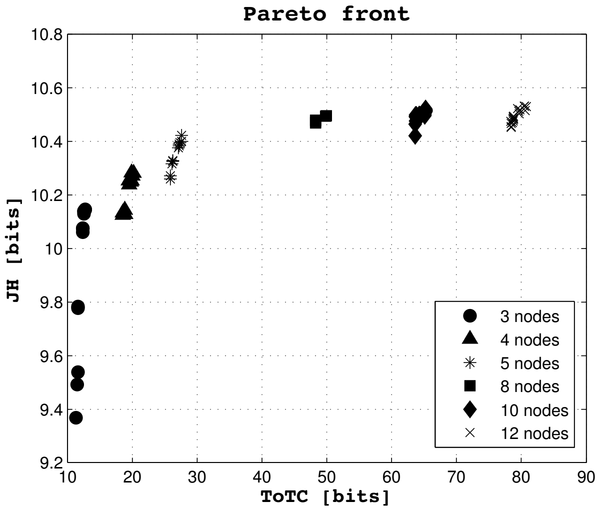

Fig. 2. Pareto front of MOOP optimal solutions considering different numbers (N) of nodes.

the Mean Square Error (MSE) indexes:

$$MAE = \frac{1}{M}\sum_{j=1}^{M} v_j, \tag{5}$$

and

$$MSE = \frac{1}{n}\sum_{j=1}^{n} u_j, \tag{6}$$

where

$$\begin{aligned} v_j &= \frac{1}{n}\sum_{i=1}^{n} |a_i|, \\ u_j &= \frac{1}{n}\sum_{i=1}^{n} a_i^2, \end{aligned} \tag{7}$$

and $a_i = (Xsim_i - Xobs_i)$, M is the number of nodes for which pressures have been simulated with the ANN. Specifically $M = (47 - N)$, $Xsim_i$ are the hydraulic pressure values that are simulated with the ANN and $Xobs_i$ are the values synthetically obtained from the hydraulic model, n is the length of each X_i.

MAE and MSE values are computed for each scenario depicted in the Pareto front. Nodes providing the least error values are considered. Minimum error values are reported in Table 2. It is interesting to notice that error values decrease rapidly when passing from 3 to 4 nodes, Figure 3. It means that the number of sensors reduction is effective. When more than 4 nodes are used, the error values weakly decreases, meaning that it is not necessary to place many pressure sensors in the WDS, once that the most informative ones are already placed. This behaviour is also confirmed in terms of information, indeed, groups formed by a number equal or higher than 4 sensors have a joint entropy almost equal to the maximum value reachable by JH.

To better display the errors between simulated and observed water pressure, these values are plotted for two nodes of the network in Figure 6.

Optimal solutions for 3, 4 and 5 sensors are highlighted with red circles on the Pareto front in Figure 4, their layout is shown in Figure 5. It is interesting to notice that sensors layout tends to cover the whole WDS, therefore they are placed following the cross shaped WDS, at least one in each main branch.

Table 2. Mean Square Error (MSE) and Mean Absolute Error (MAE) between observed and simulated water pressure generated by the ANN providing as input different groups of N nodes.

N	$MSE[m^2]$	$MAE[m]$
3	0.0036	0.0374
4	0.0025	0.0306
5	0.0016	0.0218
8	0.0014	0.0176
10	0.0012	0.0173
12	0.0011	0.0170
15	0.0010	0.0174
17	0.0010	0.0162

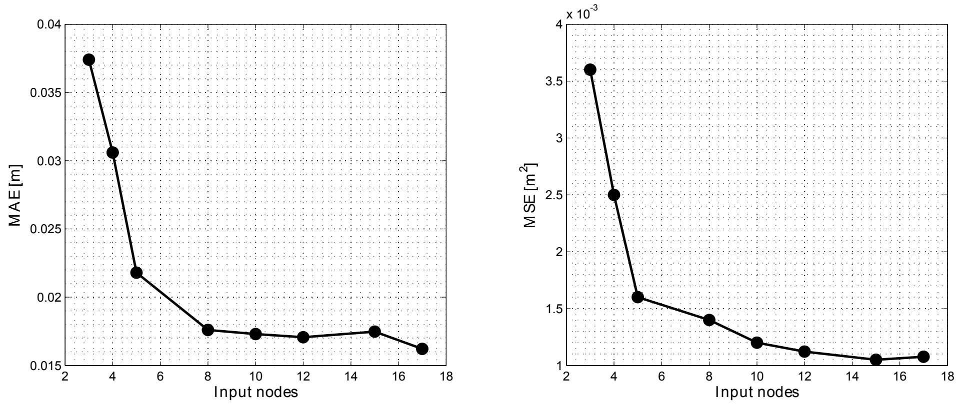

Fig. 3. MSE and MAE values against the number of nodes used to generate pressure values in all the other nodes of the WDS.

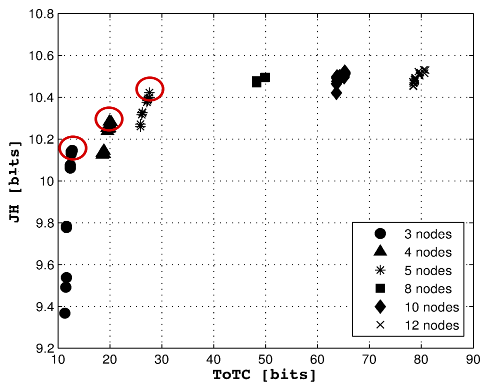

Fig. 4. Pareto front of optimal solutions, red circles highlight the group of sensors with minimum error which location is represented in Figure 5.

6. Conclusions

In this work an Artificial Neural Network is used to simulate the water pressure distribution at every node of a water distribution system. The input values are the water pressure values at a few nodes that are considered the most representative of the distribution system. They have been chosen through an entropy-based approach that allows to chose nodes depending on the amount of information that they provide about the water pressure.

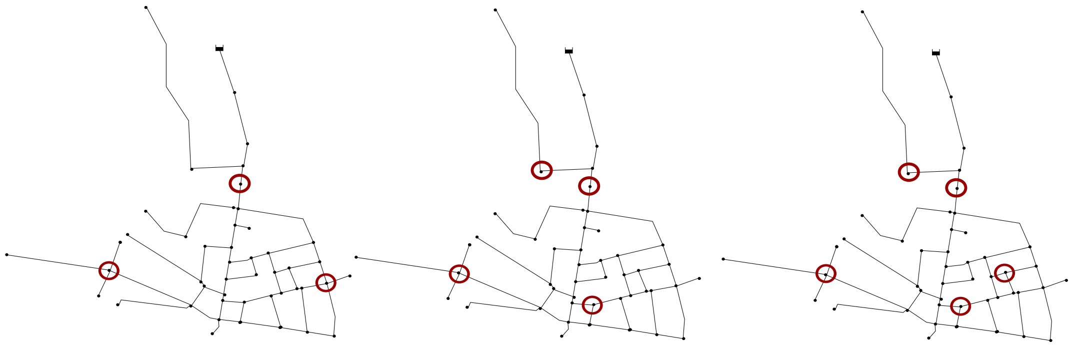

Fig. 5. Piedimonte WDS and three configuration of sensors ranging from 3 to 5.

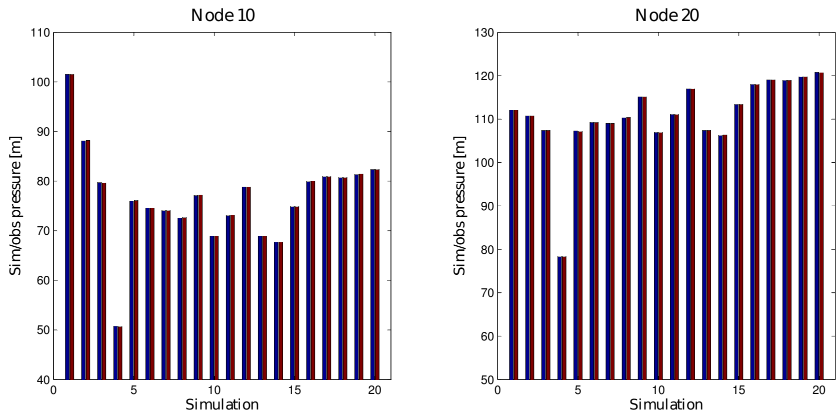

Fig. 6. Observed (red) and simulated (blue) water pressure in two nodes (10 and 20) of the WDS.

The analysis of the water distribution network shown that it is possible to deploy a small number of pressure monitoring nodes. Then, from the pressure data series that they register, the pressure values in all the other nodes of the network are evaluated. The methodology is useful when it is necessary to reduce the number of pressure monitoring sensors: from a redundant number it is possible to determine which ones can be removed from the WDS without any loss of information.

Simulated water pressure distributions are compared to observed ones to determine the goodness of the approach. Since the minimum error between simulated and observed water pressure at each node presents small value, the methodology has interesting future perspectives.

References

[1] Z. Kapelan, D. Savic, G. Walters, Optimal sampling design methodologies for water distribution model calibration, Journal of Hydraulic Engineering 131 (3) (2005) 190200.

[2] R. Meier, B. Barkdoll, Sampling design for network model calibration using genetic algorithms, Journal of Water Resources Planning and Management, ASCE 126 (4) (2000) 245250.

[3] J. Vitkovsky, J. Liggett, A. Simpson, M. Lambert, Optimal measurement site locations for inverse transient analysis in pipe networks, Journal of Water Resources Planning and Management, ASCE 129 (6) (2003) 480492.

[4] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423, 623–656.

[5] P. F. Krstanovic, V. P. Singh, Evaluation of rainfall networks using entropy: I. theoretical development, Water Resources Management 6 (1992) 279–293. 10.1007/BF00872281.

[6] Y. Mogheir, V. P. Singh, Application of information theory to groundwater quality monitoring networks, Water Resources Management 16 (2002) 37–49. 10.1023/A:1015511811686.

[7] C. Yoo, K. Jung, J. Lee, Evaluation of rain gauge network using entropy theory: Comparison of mixed and continuous distribution function applications, Journal of Hydrologic Engineering 13 (2008) 226–235.

[8] L. Alfonso, A. Jonoski, D. Solomatine, Multiobjective optimization of operational responses for contaminant flushing in water distribution networks, Journal of Water Resources Planning and Management, ASCE 136 (1) (2010) 48–58.
[9] L. Alfonso, A. Lobbrecht, R. Price, Optimization of water level monitoring network in polder systems using information theory, Water Resources Research 46(12) (2010) W12553.
[10] L. Alfonso, E. Ridolfi, S. Gaytan-Aguilar, F. Napolitano, Russo, F., Ensemble entropy for monitoring network design, Entropy 16 (2014) 1365–1375.
[11] E. Ridolfi, L. Alfonso, G. D. Baldassarre, F. Dottori, F. Russo, F. Napolitano, An entropy approach for the optimization of cross-section spacing for river modelling, Hydrological Sciences Journal 59 (1) (2013) 126137.
[12] W. de Schaetzen, G. Walters, D. Savic, Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms, Urban Water 2 (2000) 141152.
[13] K. Behzadian, Z. Kapelan, D. Savic, A. Ardeshir, Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks, Environmental Modelling & Software 24 (2009) 530–541.
[14] G. Yoo Do, D. Chang, H. Jun, J. Kim, Optimization of pressure gauge locations for water distribution systems using entropy theory, Environ Monit Assess. 184(12) (2012) 7309–22.
[15] L. Rossman, EPANET2 User's Manual., Water Supply and Water Resources Division, National Risk Management Research Laboratory, Cincinnati, OH45268, 2002.
[16] R. Abrahart, P. Kneale, L. See, Neural networks for hydrological modelling, Taylor & Francis, London (UK), 2004.
[17] R. Magini, I. Pallavicini, R. Guercio, Spatial and temporal scaling properties of water demand, Journal of Water Resources Planning and Management 134 (3) (2008) 276–284.
[18] I. Vertommen, R. Magini, M. da Conceio Cunha, Demand uncertainty in modelling water distribution systems, WIT Transactions on the Built Environment 122 (2011) 3–13.
[19] E. Ridolfi, I. Vertommen, R. Magini, Joint probabilities of demands on a water distribution network: A non-parametric approach, 11th International Conference of Numerical Analysis and Applied Mathematics 2013, ICNAAM 2013 1558 (2013) 1681–1684.
[20] A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw-Hill. New York, 3rd ed., 1991.
[21] I. Pallavicini and R. Magini, Experimental analysis of residential water demand data: probabilistic estimation of peak coefficients at small time scales, Proc. CCWI2007 & SUWM2007 Conf. Water Management Challenges in Global Change, UK, 2007.
[22] I. Vertommen, R. Magini, M. C. Cunha, Water demand uncertainty: the scaling laws approach, Water Supply Systems Analysis: Selected topics, 2012.
[23] M. B. Fiering, Multivariate Technique for Synthetic Hydrology Journal of Hydraulics Division, Proceedings of the American Society of Civil Engineers, pp. 43-60, 1964.
[24] I. Vertommen, R. Magini, M. da Conceio Cunha, Scaling water consuption statistics, Journal of Water Resources Planning and Management 04014072 (2014).
[25] A. Kraskov, H. Stoegbauer, R. G. Andrzejak, P. Grassberger, Hierarchical clustering based on mutual information, Europhysics Letters 70(2) (2005) 278.