

WDNs Calibration Using k-means Algorithm for Pipes Clustering and a Hybrid Model for Optimization

Rodrigo Freitas¹, Bruno Melo Brentan², Gustavo Meirelles Lima³, Edevar Luvizotto Junior⁴

^{1,2,3,4}Laboratory of Computational Hydraulics,
University of Campinas, Campinas - SP, Brazil.

¹rodrigo-freitas@hotmail.com.br, ²brunocivil08@gmail.com, ³limameirelles@gmail.com,
⁴edevar@fec.unicamp.br.

ABSTRACT

The modelling of Water Distribution Networks (WDNs) is an important issue for an efficient operation of water systems. The hydraulic simulations are used to establish the optimal conditions for pumps and valves, helping stakeholders to manage the systems. A reliable model should guarantee the minimal uncertain becoming the simulations the most real as possible. The calibration processes is used to reduce the uncertainties associated to pipe roughness and nodal demand, pipe status or leakage flow. Due to the high degrees of freedom surrounding this process, the improvement of available information of the hydraulic state of the network and the reduction of parameters to be calibrated can conduct for more reliable calibrated model. In this sense, this work present a hybrid calibration method, consisting in three stages defined as grouping pipes with k-means algorithms, pressure estimation with artificial neural networks (ANN) and an optimization process using particle swarm optimization (PSO) algorithms. A comparison between the classical approaches for roughness calibration is presented, reinforcing the improvement of the calibration process through the uncertain reduction.

Keywords: Water distribution networks, roughness calibration, k-means, ANN, particle swarm optimization

1. INTRODUCTION

The calibration process is an important step during the water distribution networks (WDN) modelling and is essential to become the hydraulic model reliable and a good representation of the reality. Computers and mathematical methods help in this process by solving more complex systems faster and generically. Recently the use of bio-inspired algorithms in the WDN calibration process is a tendency participating for example as an optimization process [1].

However, the calibration process needs reliable parameters to ensure the accuracy of the model, that means the necessity of a great number of monitored nodes. Another problem is the increase of the computational time associated with the complexity of the system [2]. The uncertain related with the calibration process can be read by two sides. In one hand the large number of variable and on the other hand the low number of measured data. These two points become the calibration problem highly undetermined, allowing multiple solutions. In order to reduce the degree of freedom, this

work presents a methodology that work in the two uncertain sources, becoming the problem more determined.

The usual approach for calibration is adjusting the open parameters in order to find the model pressure as similar as possible to the monitored pressure representative nodes. In this hybrid method representative nodes are chosen to be monitored while the other nodal pressures are estimated with the aid of a trained artificial neural network (ANN). With the pressure at all nodes, the network is clustered using similar pipe features (diameter, age, material) aided by a *k*-mean algorithm. The roughness of each group is finally calibrated and optimized with a PSO algorithm separately. With this method we achieve better results when compared with the traditional approach of roughness calibration (many variables, few measurement data), because are able to reduce the degree of freedom regarding the hydraulic and physical features of the problem.

2. METHODS

This hybrid method proposed at this work consists in three specific steps defined as: (1) pipe clustering, (2) nodal pressure forecasting and (3) roughness calibration. In the first step the network is segregated in clusters to reduce the number adjustable roughness. In the second step an ANN is used to amplify the number of known network information by forecasting the nodal pressure in non-monitored nodes. Finally, with all nodal pressures known and pipe clusters, PSO calibration process adjust the roughness of each group.

2.1. Artificial Neural Network

Artificial Neural Networks are bio algorithms inspired in the human's neural system. As its bio inspiration the algorithm is composed by artificial neurons which receives multiple information, processes it and then transmits one to the next neuron in sequence.

The ANN used in this work is a multi-layer perceptron (MLP) type, composed by hidden neuron layers to process the data (figure 1). In hydraulic engineer modeling the MLP are widely used as a tool to forecast and estimate the demand in WDN [3], reservoir level [4] and others prominent authors.

The MLP trained to estimate the nodal pressure in the network uses the information of monitored nodes to determine the pressure alt all other non-monitored nodes. The effect of the MLP using in the calibration process is the reduction of the uncertain, helping to improve the quality of the final model. For this work, the MLP is performed with 2 hidden layers, each one with 20 neurons.

2.2. Sensibility Analysis

To choose the best nodes to monitor, a sensibility analysis is made. As represented by the equation (1) it is a comparison between the node pressure's variation and for a demand variation. The higher the accumulated sensibility the better this node is to be monitored.

$$S_{i,j} = \left| \frac{\delta H_j}{\delta Q_{Di}} \right| = \left| \frac{H_j(Q_{Di}) - H_j(Q^*_{Di})}{Q_{Di} - Q^*_{Di}} \right| \tag{1}$$

where $H_j(Q_{Di})$ is the hydraulic head simulated in node j with a demand Q_{Di} in the node i and $H_j(Q^*_{Di})$ is the hydraulic head after a demand's variation in node i from Q_{Di} to Q^*_{Di} ;

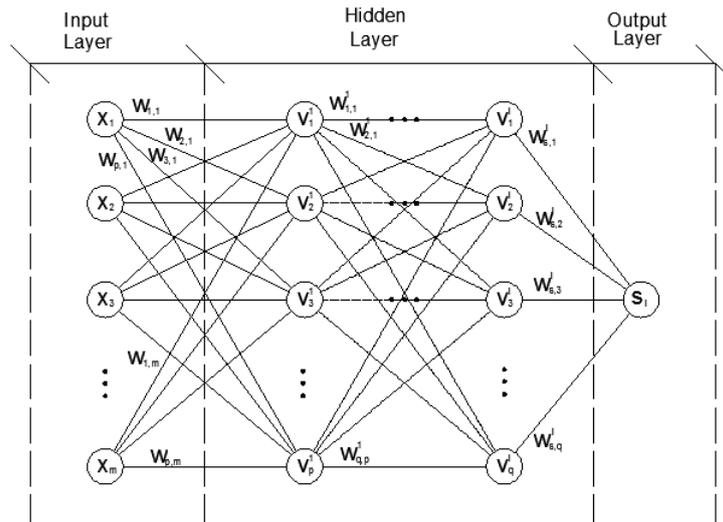


Figure 1: Generic Multi-Layer Perceptron;

2.3. *k*-means algorithm

The *k*-means is a non-supervised learning clustering algorithm based on the distance between a centroid and the data. Firstly it creates random points in the space, which the quantity matches the amount of desired clusters, defining the centroids [5]. After that the data is separated according to the distance to the centroids, usually the Euclidean distance. With the grouped data, the previous centroids move to be the centroid of each cluster and the distance between the centroids and each data point is once again clustered and the centroid relocated [6]. This iterative process is repeated until a maximum number of iterations or until the centroids' movements are insignificant.

The aim of the *K*-means algorithm in this work is to divide all the network nodes' into clusters with hydraulic similarities aiming with that the simplification of the calibration process. The main characteristic used to create the clusters was the pipes age, material and diameter.

2.4.PSO Algorithm

The PSO, proposed by [4] is a bio-inspired algorithm based on birds and fishes. The algorithm starts with a random set of solutions (initial swarm) which moves towards the optimal solution through iterations guided by information shared among all the swarm elements. Each element has a velocity and a position which changes after each iteration.

Initially its characteristics (velocity and position) are randomly generated. In the next step velocity and the particle position updated according to the performance of each particle, as shown in (5) and (6).

$$v_{k+1}^i = wv_k^i + c_1 \text{rand} \frac{(p^i - x_k^i)}{\Delta t} + c_2 \text{rand} \frac{(p_k^g - x_k^i)}{\Delta t} \quad (5)$$

Where w is the weight factor, c_1 is the cognitive factor and c_2 is the social factor. The suggested values for each parameter can be found at [7].

$$x_{k+1}^i = x_k^i + v_{k+1}^i \Delta t \quad (6)$$

3. RESULTS

3.1. Case Study

The studied network is known as Modena networks [8] is composed by 317 pipes, 268 nodes and it will be adapted for the calibration study with an age zones creation. It was supposed a radius growth with the center older and the periphery newer. Each sector was considered with different material tubes and so on different values for the Hazen-Williams coefficient. The older region, with used cast iron pipes presents a Hazen Williams Coefficient equal to 90 while the intermediate and the newer region presents this coefficient equal to 130 (new cast iron pipes) and 150 (P.V.C pipe) respectively. Figure 3 shows the network, its elements and the three regions proposed.

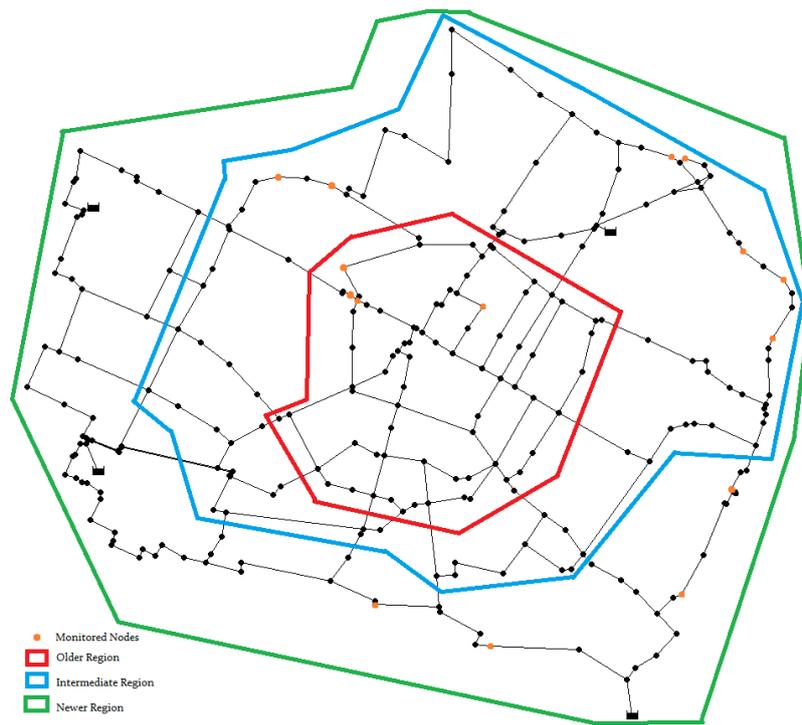


Figure 2: Monitored nodes and Network separated by age and sectors;

3.2. Estimated pressure by ANN

With the help of the sensibility analysis shown in item 2.2, the 5% more sensible nodes were chosen to be monitored and represent the network (figure 3). With the monitored nodes, the pressures in all the other nodes were predicted by an ANN.

The Figure 4 shows the minimal, average and maximal pressure errors in all nodes. It is possible to identify the high accuracy of the ANN performance, allowing a good hydraulic state estimation for the calibration process. Figure 5 shows the average node pressure error surface. With it is possible to locate the most problematic nodes (nodes with higher errors).

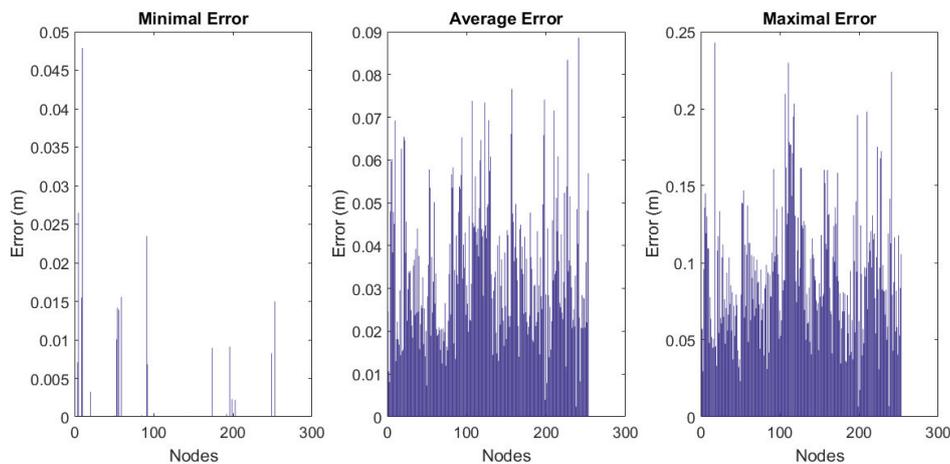


Figure 3: Graphic bars showing the minimal, average and maximum pressure error in all nodes;

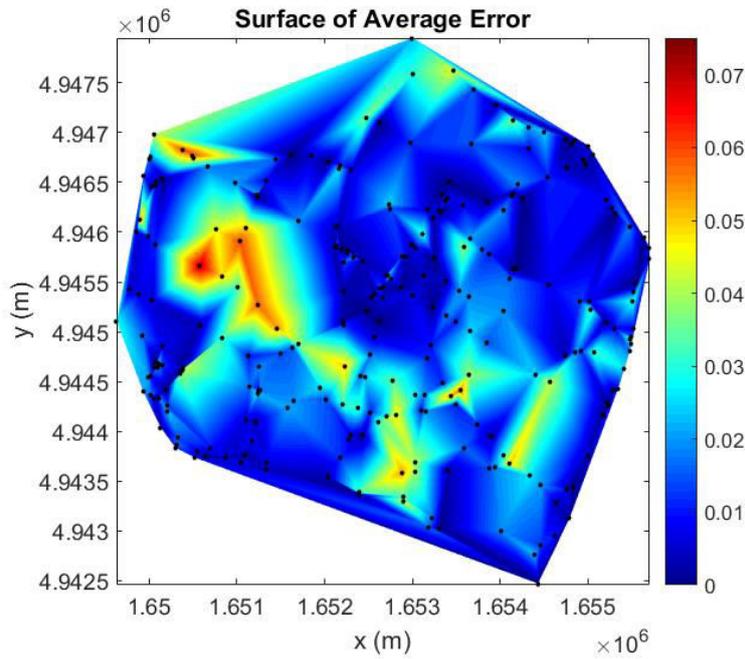


Figure 4: Surface of average error in the network

3.3. Hazen-Williams’ coefficient calibration with clustering and PSO

During the clustering process the main features used to create the groups are the pipes age, material and diameter. With the pipe clusters defined by the k-means algorithm, the PSO optimization is run to calibrate the roughness of each group. In order to compare the effect of each action to reduce the freedom degree of the calibration problem, a comparison among the classical approach (using only the monitored pressure to calibrate the roughness of each pipe), the state estimation process (using the monitored pressure to estimate the pressure at all nodes) and the hybrid methodology presented at this work is done.

Table 1 is a comparative table where the average and maximum error of the pressure at all nodes, the error of the flow in all pipes, the Hazen-Williams Coefficient error and the error of the monitored pressure are shown.

Disregarding the monitored pressure, the maximum and average errors were considerably lower in the proposed hybrid method when compared to the other two methods. The average error in the Hazen-Williams coefficient (calibration parameter) was 27.98 which is approximately 37% lower than the error in the two other calibration methods.

Looking at the pressure errors, it is possible to observe that although the errors in the monitored nodes are larger than in the other calibration methods, when the error is analyzed in all the nodes the errors, both medium and maximum, are the smallest in the three methods with the average error 33% and 68% and the maximum error 03% and 71% smaller when compared with the ANN Calibration and the classical calibration respectively.

Table 1. Comparison between each calibration method for hydraulic parameters

	Pressure at all nodes (m)	Flow in all pipes (l/s)	C	Monitored Pressure (m)
Clustering+ANN Calibration				
Average	0.82	1.31	27.98	0.90
Maximum	4.60	9.63	108.00	2.60
ANN Calibration				
Average	1.22	2.38	44.35	0.63
Maximum	4.75	23.06	125.78	3.60
Classical Calibration				
Average	2.54	2.45	44.75	0.01
Maximum	16.06	21.50	124.10	0.08

4. CONCLUSIONS

The calibration process of pipe roughness is an important step during the development of hydraulic models. The high degree of freedom presents in the calibration problems become it hard to be solved with a conclusive solution. The methodology presented at this work intends to propose a new alternative to the calibration process aided by algorithms and mathematical-computational methods aiming an optimization of the calibration process which the principal advantage over other methods is the reduction of uncertainties through two different fronts. While the ANN increases the amount of information available by predicting the pressure in all the system's nodes the clustering reduce the amount of parameters to be calibrated without losing quality in the process, sectoring the network . With this method it was possible to obtain errors from 30% to 70% lower than the mentioned methods popularly used.

The results obtained were good and promising. This methodology proved to be an interesting approach to the calibration problem deserving more attention, studies and applications to consolidate it, which we will continue to do.

References

- [1] B. M. Brentan, E. Luvizotto Júnior. "PSO applied to reduce the cost of energy in water supply networks". *Applied Mechanics and Materials*, 409–410, 703–706, 2013.
- [2] AFSHAR, A.; HAMIDEH, K.; SAADATPOUR, M. Particle Swarm Optimization for Automatic Calibration of Large Scale Water Quality Model (CE-QUAL-W2): Application to Karkheh Reservoir, Iran. *Water Resour Manage*, v. 25, n. 10, pp. 2613–2632, 2011.
- [3] CORDOBA, G. A. C.; TUHOVCÁKA, L.; TAUSA, M. Using artificial neural network models to assess water quality in water distribution networks. In: 12th INTERNATIONAL CONFERENCE ON COMPUTING AND CONTROL FOR THE WATER INDUSTRY, Perugia, Itália, 2014.
- [4] KENNEDY, J.; EBERHART, R. Particle Swarm Optimization. *Evolutionary Computation – IEEE*, pp. 1942–1948, 1995.
- [5] HARTIGAN, John A.; SONS, John Wiley &. *Clustering Algorithms*. Department Of Statistics Yale University: Wiley, 1975. 369 p.
- [6] HARTIGAN, J. A.; WONG, M. A.. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal Of The Royal Statistical Society. Series C (applied Statistics)*. New Haven, Connecticut, p. 100-108. jan. 1979.
- [7] HASSAN, Rania et al. A Comparison of Particle Swarm Optimization and the Genetic Algorithm. 46th Aiaa/asme/asce/ahs/asc Structures, Structural Dynamics And Materials Conference, [s.l.], p.1-13, 18 abr. 2005. American Institute of Aeronautics and Astronautics. <http://dx.doi.org/10.2514/6.2005-1897>.
- [8] Bragalli, C., D'Ambrosio, C., Lee, J., Lodi, A., Toth, P., 2012. On the Optimal Design of Water Distribution Networks: a Practical MINLP Approach. *Optim. Eng.* 13, 219-246.