16th Conference on Water Distribution System Analysis, WDSA 2014

# Use of Metamodels in Real-Time Operation of Water Distribution Systems

F. Keizo Odan[a]*, L. F. Ribeiro Reis[b], Z. Kapelan[c]

*[a]Federal Center of Technological Education, Belo Horizonte, Minas Gerais (Brazil)*
*[b]University of São Paulo, São Carlos, São Paulo (Brazil)*
*[c]University of Exeter, Exeter, Devon, (United Kingdom)*

**Abstract**

This paper presents a methodology for real-time pump scheduling in water distribution system. The methodology uses an optimizer linked to other two modules: one to predict the hydraulic behavior of the system and another to forecast the demands. A metamodel is used instead of a traditional hydraulic simulator to speed up relevant computations, i.e. evaluations of large number of candidate pump schedules generated during the optimization process. The methodology presented is applied to a real-life water distribution system in Brazil. The results obtained show that the use of metamodel can generate low cost pump schedules in a computationally fast manner.

*Keywords:* Optimization; real-time operation; metamodelling; water distribution systems.

## 1. Introduction

Water distributions networks are complex systems composed of pipes, reservoirs and pumping stations. These systems are subject to demands from consumers of many types as result of diverse needs. Moreover, the water companies must meet water demand with proper quantity and quality, while trying to reduce cost through an efficient operation.

The energy used at pumping stations is one of the major operational costs. In order to save energy and reduce cost, the operators can have their decision supported by the use of mathematical model to rationalize the process and obtain

---

* Corresponding author.
  *E-mail address:* frederico_keizo@yahoo.com.br

optimized strategies to operate the water system, through the real-time optimization of operation, i.e., produce the optimized operation for the next 24 hours every new period, for example, every hour, using constantly updated data, to account for the variation of demand and other data throughout the day. The pump scheduling optimization can be performed by coupling to the optimizer a simulation model, to reproduce the behavior of the water system, and a demand predictor, to forecast the water demand.

Recently, population based algorithms, like Genetic Algorithms (GAs) have been used to optimize the pump schedules ([1]; [2]; [3]) due to their capability to perform an effective exploration of the search space and hence ability to reach a global optimum. However, population based techniques requires a great number of function evaluations, making these methods computationally inefficient, and therefore impractical for real-time operation. Alternative, hybrid type approaches have been developed recently ([11]) but typically require introduction of some simplifications and/or approximations (e.g. linearization of fundamental hydraulic equations).

Some authors proposed the use of Artificial Neural Network (ANN) ([4]; [5]) to substitute the traditional simulation model like EPANET2 ([6]) to speed up the process. This paper aims to take a step further by using new promising techniques. The methodology presented here is implemented in a real-time framework with novel post-processing of schedules identified by the optimization to ensure feasibility of schedules that are actually implemented (in terms of pump switches)[10]. In addition, the methodology proposed here makes use of adaptive ANN training ([7]) as an alternative to a more standard Back Propagation method and, makes use of a multi-method optimization algorithm ([8]) as an alternative to a more commonly used GA.

This article was organized in order to summarize the methodology used and the case study as well as the main results, followed by conclusions.

## 2. Methodology

### 2.1. Real-time Pump Scheduling Framework

The real-time pump scheduling framework is shown in Fig.1. As it can be seen from this figure, real-time pump schedules are generated by using the pump scheduling system (i.e., software tool based on the methodology described in this paper) which is run every hour. The real-time pump scheduling procedure is as follows:

1) Receive latest data from the Data Acquisition part of the SCADA (Supervisory Control and Data Acquisition) system. These data comprise tanks levels, pump and valve statuses and flows in/out of the system that are used to estimate the current water consumption in the system.
2) Forecast water demands for the next 24 hours by using current and past water consumption data;
3) Update the Hydraulic Simulation Model (or Simulator) of the analyzed WDS by using data obtained in steps 1 and 2;
4) Run the Optimization Model to identify the optimized system schedule, i.e., best pump schedule for the next 24 hours. During the optimization process use the Hydraulic Simulator to evaluate alternative pump schedules generated by the optimizer in terms of operational cost and to check the feasibility of the generated pump schedule (with respect to constraints mentioned in section 2.3.1);
5) Implement the optimal pump schedule identified in the previous step for the next hour only. This is done via the Supervisory Control part of the SCADA system (see Fig. 1).
6) Repeat steps 1 to 5 continuously, i.e., until scheduling is desired.

### 2.2. Water Demand Forecasting Model

Implementing a water demand forecasting model involves selecting a model, defining its best structures and input variables, as well as defining its evaluation criteria. In this study it was chosen to apply the Hybrid Dynamic Neural Network (DAN2-H) developed by [9]. It is a self-constructing neural network which models the previous water consumption with the forecasted demand produced by a Fourier Series.
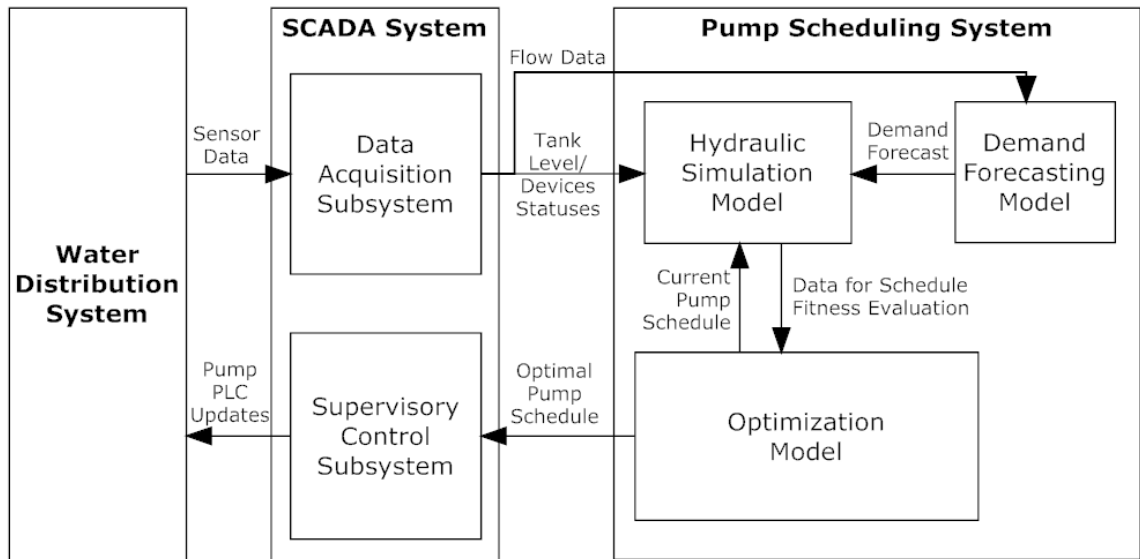
Fig. 1- Pump Scheduling Framework

The demand forecasting model estimate 24hours ahead at each time step (1hour), whose values are used to perform the operation of a water network for the next 24 hours. Table 1 presents the inputs and outputs of the model. The character "C" denotes the consumption and FS correspond to the forecasted demand by Fourier Series. The forecasting model uses both data of the previous 4 hours plus the data lagged by 168 hour. The model is run 24 times in a recursive fashion to obtain demands for the next 24 hours. Further details can be found in [9].

Table 1 – DAN2-H model and its respective input and output

| Input | Output |
|---|---|
| C(t), C(t-1), C(t-2), C(t-3), C(t-168), FS(t), FS(t-1), FS(t-2), FS(t-3), FS(t-168) | C (t+1) |

### 2.3. Optimization Model

The optimal pump schedules for the next 24 hours is determined by optimization model given the forecasted demands for the same time period. The pump scheduling problem is formulated and solved as an optimization problem with specific objectives, constraints and decision variables.

### 2.3.1. Optimisation Problem

The objective function minimises the cost of energy used by the pumps over the 24 hour scheduling horizon. The following constraints were considered in optimization problem:

- Pressure was not allowed below a pre-specified minimum threshold. This work used the threshold of 10 m, based on Brazilian standards;
- Final water tank levels (i.e. at the end of the scheduling horizon) must not be lower than the corresponding tnak levels at the beginning of the same time horizon. If it occurs, it will lead to unsustainable longer-term operation of the system;

- Number of pump switches must be lower than a pre-specified threshold. Excessive pump switching causes its premature wear and tear and leads to increased maintenance costs. A threshold of four pump switches was considered here.
- Occurrence of error during hydraulic simulation, caused by system been disconnected or occurrence of negative pressure. Those errors were not allowed and the corresponding infeasible solutions were discarded.

In order to reduce such search space, compared to binary representation, it was decided to use the representation proposed by [3], called relative time-controlled trigger. This representation allows controlling the pump through time and explicitly limiting the number of pump switches. The control is represented by successive pairs of decision variable $t_i$ and $t'_i$, representing the period the pump should be OFF and ON, respectively. In this paper three (3) pairs were used, which means the pump switch was limited to three (3).

### 2.3.2. Hydraulic Simulation Model

The hydraulic simulation is performed here by using a metamodel based on ANN rather than a conventional hydraulic simulation model. The single hidden layer neural network is calibrated using the Adaptive Merging and Growing Algorithm (AMGA) ([7]). This algorithm merges and adds neurons of the hidden layer based on the progress and learning of the neurons of that same layer and, at the same time, avoids the trial and error process to determine the appropriate number of neurons of the hidden layer, which is necessary in the traditional back propagation neural network.

The Metamodel I was developed following the methodology proposed by [5]. The development consists of generating data using EPANET 2 to calibrate the ANN, considering the complexity of the water system and its operation. The metamodel built uses as an input the "Total Demand", the number of "Pumps (turned) ON" and the "Initial Reservoir (tank) Level". The "Total Demand" is represented by $N_{demand}$ inputs, one for each District Metering Area (DMA), the "Pumps ON" is represented by $N_{pumps}$, one for each group of pumps and the "Initial Tank Level" is represented by $N_{tank}$ inputs, totalizing ($N_{demand} + N_{pumps} + N_{tank}$) inputs as illustrated in
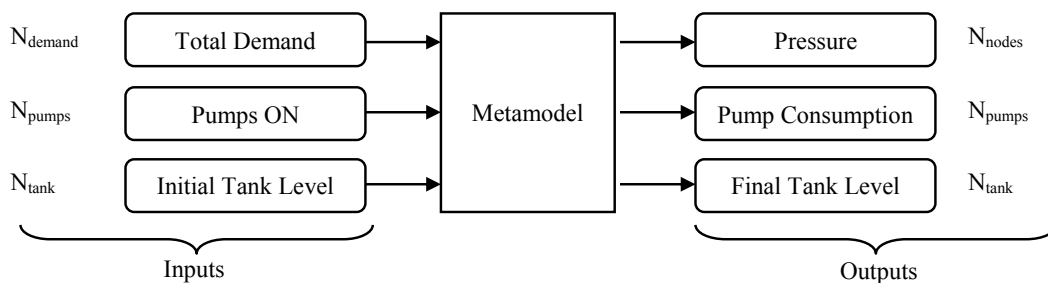


Fig. 2 – ANN based Metamodel I

The metamodel predicts the total energy consumed by each group of pumps ($N_{pumps}$), the final level at each reservoir ($N_{tank}$) and the "Pressure" at each monitored critical node ($N_{nodes}$), totalizing ($N_{nodes} + N_{pumps} + N_{tank}$) outputs. The metamodelling process used only feasible pump schedules to calibrate the metamodel. So if an infeasible pump schedule, i.e., that violates the constraints, is evaluated by the metamodel, it will produce an output regardless of its infeasibility. To overcome this problem, another metamodel was created only to classify the system pump schedule as either feasible or infeasible, to distinguish between the two during the optimization process. This was called Metamodel II. It uses the same input as Metamodel I, but it predicts only one output – if the pump schedule was infeasible or not.

*2.3.3. Optimization Method*

A Multi ALgorithm Genetically Adaptive Method (AMALGAM), proposed by [8], was chosen as optimization method to solve the pump scheduling problem. AMALGAM simultaneously uses $k$ (=4) different search algorithms, which learns from each other by sharing data from a common population of points. Initially it creates a random population of $N$ (=100) solutions, then each algorithm generates an offspring solutions proportional to their success in the previous generations of solutions. The main idea of the method is to combine the strengths of different metaheuristics to have a more efficient and effective optimization process.  It was used the same metaheuristics as in the paper of [8]: the NSGA-II (Non-dominated Sorting Genetic Algorithm II), the PSO (Particle Swarm Optimization), the Adaptive Metropolis Search (AMS) and the Differential Evolution (DE).The optimization was limited to 2,500 objective function evaluations, with a population of 100 solutions, meaning that 25 iterations were used. The parameter values adopted to implement the methods can be consulted at [8].

The real-time optimization of the system operation is performed to each hour of the day, i.e., 24 times per day. Each time optimized pump schedule for the next 24 hours is generated. However, only the pump schedule for the first hour is actually implemented. Although the optimized pump schedules meet the optimization constraints (defined in previous section) for the forthcoming 24 hours, this may not be true for the actual schedule implemented, i.e., the actual 24-hour schedule may exceed the maximum number of pump switches allowed. In order to help overcome this problem, the optimized pump schedule of next 24 hours are post-processed in a novel way, before implementing the first hour schedule, to minimize pump switch and the pump use during peak hours, i.e., period when energy is more expensive. Further details can be found in Odan [10]. It is important to notice that the post-processing technique only helps the mentioned problem, i.e., the quality of optimized schedule are ruled by the optimization method.

The metamodel is used to evaluate the pump schedule generated by the optimizer. However, as the metamodel is a surrogate of the traditional simulation model, the produced objective functions are only approximations as well. So it is necessary to reevaluate the final solutions using the EPANET 2 to obtain their real objective function values. Usually after the reevaluation, it is noticed the solutions are near the Pareto Front, but often not part of it. So, to ensure better solutions, these solutions are once again optimized by performing additional 500 objective functions evaluations.

## 3. Case Study

A DMA of a real system operated by the Autonomous Department of Water and Sewage (DAAE), from the city of Araraquara, state of São Paulo, Brazil, supplied from the underground wells was used here. Its hydraulic model was previously calibrated. Although the well levels vary with time, an average value was used for the selected month of analysis

The electricity has two different prices for the peak and off-peak hours, R\$ 5.41 and R\$ 0.76, respectively. The peak hour refers to the three consecutive hours between 18:00 and 21:00 pm. In order to allow comparison between optimized operation and the operation currently practiced by the water company, it was necessary to identify a period of time without faulty data which was done on a monthly basis. At the end, one month of data was chosen for the DMA to optimize its operation, from April 27 to May 27 of 2010. Cost was calculated in Real (R\$), where US\$1.00 was equivalent to R\$2.28, according to the currency exchange on 9th September of 2013. The DMA characterized by residential consumption has 18.5 km of pipes, which consumes 54 m³/h. The corresponding hydraulic model (Fig. 3 (a)) has 252 nodes and 282 pipes.

The Water Supply Facility (WSF) is comprised by the reservoir, the pumping station and two wells, detailed in Fig. 3 (b). The water is pumped by two wells called Well I and Well II to the reservoir R8, which supplies the DMA. Both pumps have144 hp and supply115 m³/h. The bottom and maximum water level of R8 are located at 715.59 and 721.99, respectively. It stores up to 500 m³.

The critical node monitored on the metamodel is represented by the red arrow on Fig. 3(a).
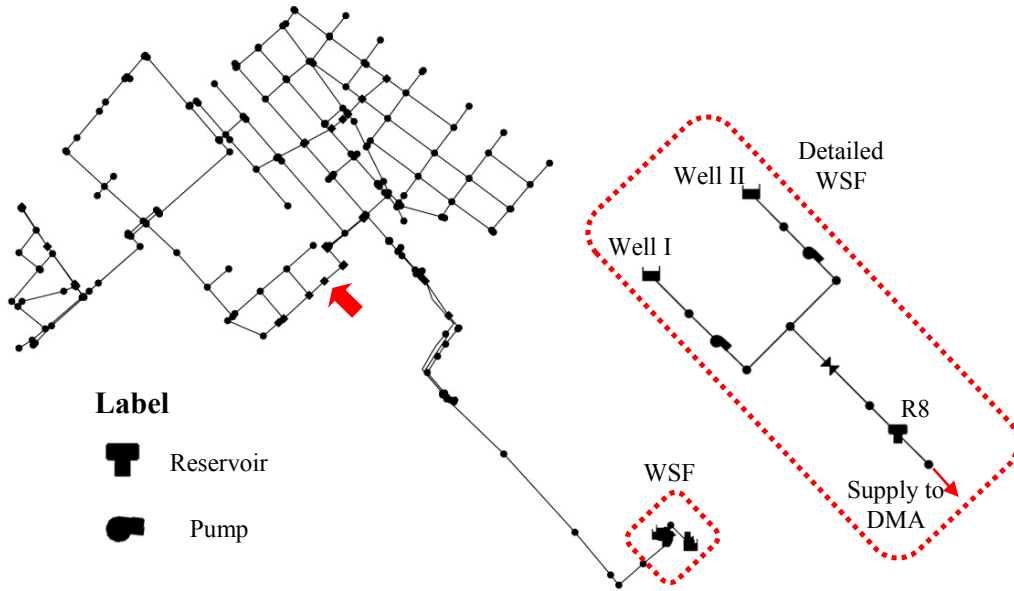
Fig. 3 – (a) DMA; (b) Detailed Water Supply Facility of DMA

It was generated 5.000 pair of input-output for the metamodelling. The calibration used 75% of the generated data, and the rest was used for the testing only. On this case study it was used 4 inputs, as there is 1 DMA ($N_{demand}$ = 1), two pumping groups ($N_{pumps}$ = 2), and 1 reservoir ($N_{tank}$ = 1). The output totalized 4 outputs: one monitored node ($N_{node}$ = 1), two pumping groups ($N_{pumps}$ = 2) and one reservoir ($N_{tank}$ = 1).

## 4. Results and Discussion

### 4.1. Metamodelling results

The results of Metamodel I calibration to predict the energy consumed by the pumps, the final reservoir level and pressure at the critical node are summarized in Table 2, using only testing data. The output was evaluated using the Coefficient of Determination ($R^2$), the Mean Absolute Error (MAE) and the Mean Percent Error (MAPE). From the results, it is observed the 38 hidden neurons metamodel built was able to predict the outputs with great accuracy.

Table 2 –Results of Metamodel I

| Variable | $R^2$ | MAE | MAPE (%) |
|---|---|---|---|
| Energy Consumed by Pump from Well I | 1.00 | 0.01 kWh | 0.22 |
| Energy Consumed by Pump from Well II | 1.00 | 0.01 kWh | 0.14 |
| Reservoir Level at R8 | 0.99 | 0.11 m | 2.39 |
| Pressure at node n29d | 0.99 | 0.12 m | 0.37 |

Regarding Metamodel II, used to predict if the pump schedule is feasible or not, it was built with 16 hidden neurons and produced a great accuracy as well, higher than 98%.

### 4.2. Pump Scheduling optimization

The methodology was applied for a period of one month to compare the optimized cost to the operation practiced by DAAE. The optimization was limited to 2,500 objective function evaluations, which proved to be good enough to produce feasible and (near) optimal solutions. The optimization took about 1.0 minutes to generate the pump schedule of the next 24h, while the optimization with EPANET 2 consumed 1.8 minutes, both performed on a desktop Intel Core i7-3770 3.40 GHz, 16 GB RAM, Windows 64 bits. The use of metamodel on the optimization process consumed 45% less processing time than using the traditional simulation model.

On Fig. 4 it was compared the optimization using the ANN (blue line with circle), the optimization using the full hydraulic model (green line with triangles) and the operation practiced by DAAE (red line). The ANN optimized solutions identified to the studied DMA were more economic for 29 of 31days, while using the full hydraulic model it was more economic in 27 days. The ANN optimized operation cost over the analyzed time period was R$ 5,818.12, the Full Hydraulic optimization cost was 5,964.03, while the DAAE operation cost was R$ 6,914.75, representing savings of about 16% and 14%. The ANN optimized pump schedule was much more expensive on 9th day than the operation practiced by DAAE, while Full Hydraulic optimization was more expensive on 2nd, 3rd, 9th and 11th days, because the optimized operation used much more energy during peak hour. On the 30th day, the difference between ANN optimized pump schedule and DAAE operation was negligible, of about R$ 6.
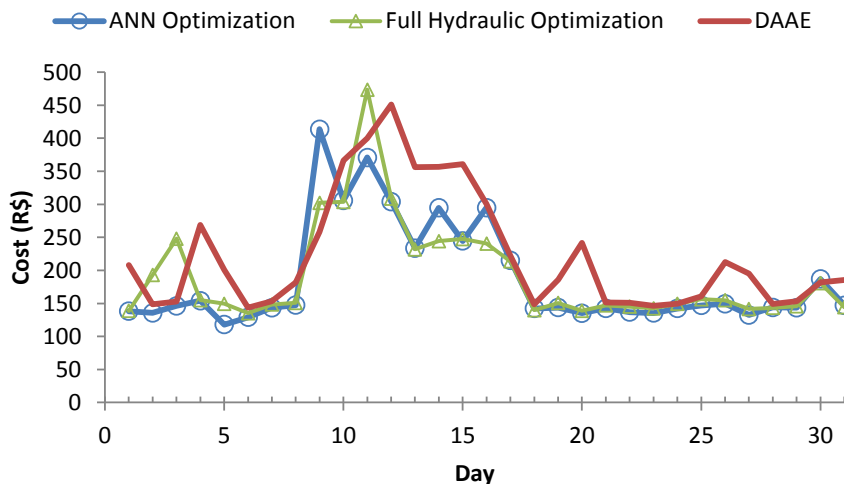


Fig. 4 - Costs of the Optimized Operation and the Operation practiced by DAAE

### 5. Conclusion

The aim of this work was to develop metrology (and software tool) for effective and efficient real-time pump scheduling by combining several methods for demand forecasting, hydraulic meta-modeling and optimization. The methodology developed was tested to an existing real-life water distribution system of Araraquara city, state of São Paulo, Brazil.

Based on the case study results obtained, the coupling of AMALGAM optimization method and the developed ANN based hydraulic metamodels can be considered a robust and efficient tool for real-time pump scheduling. The ANN and Full Hydraulic optimization approach reduced the cost of pump scheduling by approximately 16% and 14% respectively, when compared to the cost of energy used in historical DAAE operation in April-May of 2010. Besides the reduced cost, it was able to reduce the computational processing time by 45%, compared to the use of the traditional simulation model (EPANET 2).

None of the previous work considered the necessity of a second metamodel to predict the feasibility of the pump schedule. It showed to be an essential tool in the search of the global optimum. The test performed without it, omitted here for space reason, was producing bad results, as it was not possible to distinguish feasible and infeasible solutions. Usually the optimization turned off the pumps for all day long, as it didn't knew it would lead to an infeasible solution.

As pointed out by [5], it is important to reevaluate the final solutions produced by the optimization process using the traditional simulation, as these solutions will be probably near the Pareto front. Instead of performing a local search ([5]), the final solutions were reevaluated using the AMALGAM.

The future research should consider objectives other than cost, for example, reliability measures. The developed methodology should also be applied to other water systems for further testing.

## Acknowledgements

## References

[1] D.A. Savic, G.A. Walters, M. Schwab, Multiobjective Genetic Algorithms for Pump Scheduling in Water Supply, Evolutionary Computing, AISB Workshop, 1305 (1997) 227-236.
[2] P.F. Boulos, Z. Wu, C.H. Orr; M. Moore, P. Hsiung, D. Thomas. Optimal pump operation of water distribution systems using genetic algorithm, H2ONET - Users Guide.MW Software INC. 2000.
[3] M. López-Ibáñez. Operational Optimisation of Water Distribution Networks. PhD thesis, School of Engineering and the Built Environment, Edinburgh Napier University, UK, 2009.
[4] D.G. Jamieson, U. Shamir, F. Martinez, M. Franchini. Conceptual design of a generic, real-time, near-optimal control system for water distribution networks. J. Hydroinformatics 9 (2007) 3–14.
[5] D.R. Broad, H.R. Maier, and G.C. Dandy. Optimal operation of complex water distribution systems using metamodels. J. WaterResour. Plann. Manage., 136 (2010) 433–443.
[6] L.A. Rossman. EPANET 2 user's manual. U.S. Environmental Protection Agency, Cincinnati, Ohio. 2000.
[7] M. Islam, A. Sattar, F. Amin, Xin Yao, K. Murase. A New Adaptive Merging and Growing Algorithm for Designing Artificial Neural Networks. IEEE Transactions on Systems, Man, and cybernetics- Part B: Cybernetics, vol. 39, no. 3, June 2009.
[8] J.A. Vrugt, B.A. Robinson. Improved evolutionary optimization from genetically adaptive multimethod search. Proc. Natl. Acad. Sci. U.S.A., 104(3), 2007, pp. 708–711.
[9] F.K. Odan, L.F.R. Reis. Hybrid Water Demand Forecasting Model Associating Artificial Neural Network with Fourier Series. J. Water Resour. Plann. Manage. 138 (2012) 245-256.
[10] F.K. Odan. Study of Reliability applied to Real Time Optimization of Operation of Water Network Supply. (2013). 210 p. Thesis (Doctorate) – School of Engineer of São Carlos, University of São Paulo, São Carlos.
[11] C. Giacomello, Z. Kapelan, M. Nicolini. Fast Hybrid Optimisation Method for Effective Pump Scheduling. Journal of Water Resources Planning and Management (ASCE), 139 (2013) 175–183.