Multiple Linear Regression Analysis of Pumps Performance in Water Pumping Plants

Fouad LAAJINE[1,2], Mohammed MACHKOR[2], Driss MAZOUZI[1]

[1]R.N.E Laboratory, Multidisciplinary Faculty of Taza, University Sidi Mohamed Ben Abdellah, Fez, Morocco
[2]National Office of Electricity and Drinking Water Morocco

Correspondence: Driss MAZOUZI (driss.mazouzi@usmba.ac.ma)

Abstract. Energy use in water supply systems represents a consequent part of global energy consumption across all sectors. This consumption is expected to rise, due to the increasing demand and the recourse to unconventional water resources. Regarding water utilities, most of their operating costs are related to energy consumption, especially pumping systems consumption. In this context, the main objective of this study was to model accurately the energy consumption of pumping systems in order to optimize the whole water supply system, thus improving its efficiency, especially in the case of a limited renovation. For this purpose, Multiple Linear Regression was fitted to model the produced kWh/m^3 ratio according to the following parameters, active and reactive energies, the daily produced water volume, the power factor (Cosφ), and the operating time of each pump. The final model describes accurately the consumption per cubic meter produced (R^2=0.91). Therefore, this model could be considered a good estimator for the calculated ratio, which was close to the experimental one. In addition, this approach considers the system's behavior while most of the comparable studies focus on the pump scheduling problem estimator for the calculated ratio which was close to the experimental one. In addition, this approach considers the real-time-data behavior while most of the comparable studies focus on the pump scheduling problem.

Key Words: energy efficiency, linear multiple regression, pumping systems, water Supply.

1. Introduction

In the upcoming decades, water demand is forecasted to significantly increase up to 35% in 2050 (UN-Water 2020). Therefore, energy needs that are related to the water pumping and distribution processes could contribute to the expected increase in energy demand (EIA 2019). Indeed, several studies have concluded that the energy consumption related to these processes accounts for 7% of the total energy used across the globe (Coelho et al., 2014). His share is expected to get bigger due to the increasing distances between resources and populations, especially in water-scarce countries, and the growing consumption per capita due to the industrialization and improvement of living standards.

To reach the sixth Sustainable Development Goal (SDG6) that aims to generalize the access to drinking water supply, the water production cost mustn't impact its price which should stay affordable to the population. Therefore, production costs should be reduced via the control of energy consumption which is typically the largest marginal cost for the production (Helena et al., 2017).

Pumping processes consume the largest fraction of total energy (Plappally and Lienhard, 2012). The pumps consumption often presents 80% to 90% of the total energy consumption (Sarbu, 2016). However, this consumption may depend on many factors such as surface water or ground water, transport differences, flat or mountain regions, etc (Rothausen and Conway, 2011) (Plappally and Lienhard, 2012). Thus, by achieving energy efficiency improvements measures, we may reduce this consumption by 25% (Moreira, et al., 2013). In this context, few studies were interested in modeling pumping systems and evaluated the influence of parameters such as the aging of the components, which can reduce the performance of the pump by up to 12% (Durmus, et al., 2008).

Many factors impact the overall energy efficiency of a water pumping system. One important factor is the operating point which should be as close as possible to the maximum energy efficiency region (Carravetta, et al., 2020). Another factor is the age of the system, especially the pumps which can lose 1.5% of their efficiency every year (Zhou et al., 2021). Many attempts have been made to address the subject of pumping optimization using different approaches to optimize either energy consumption or the overall cost of energy. To optimize the energy consumption, a model based on ant colony was utilized (Ostfeld and Tabultzev, 2008) to provide an effective model that minimizes the design and the operating costs. This model act on parameters such as: the diameters of the pipes, the pumping station's maximum power, and the tank's storage. Another work targeting the minimization of energy consumption (Giustolisi O et al., 2012) used a multiobjective strategy to minimize losses by controlling pressure in the water distribution system and thus minimizing the energy consumption.

On the other hand, various studies have focused on optimizing the operation of the existing water systems by using deterministic methods such as Dynamic Programming (Fallside and Perry, 1975) which worked on minimizing the energy cost by load shifting using a hierarchical optimization to reduce the problem to 10 control parameters. Given the difficulty of the deterministic approach, different metaheuristic methods were developed, the widely used among them is Genetic Algorithms (Wu et al., 2014) which used evolutionary computation to maximize the efficiency of a system consisting of 2 pumps. Although these efforts employ evolutionary optimization techniques. However, in literature, particular attention has been paid to the applicability of linear regression (Puleo et al., 2014) and multiple linear regression (Adamowski et al., 2012) to the pumping scheduling optimization problem.

This work targets the pumping system of the Bab Louta drinking water production located in the province of Taza, Morocco. In this perspective, a multiple linear regression (MLR) was fitted to model the produced kWh/m^3 ratio according to the input parameters by using real-Time-data.

2. Materials and methods

Conventional water supply systems (WSS) consist of sets of structures and facilities providing products with a suitable quantity and quality for domestic and industrial use (Luna et al., 2019). To develop energy and hydraulic model, WSS systems should be evaluated in terms of mass and energy (Figure 1).

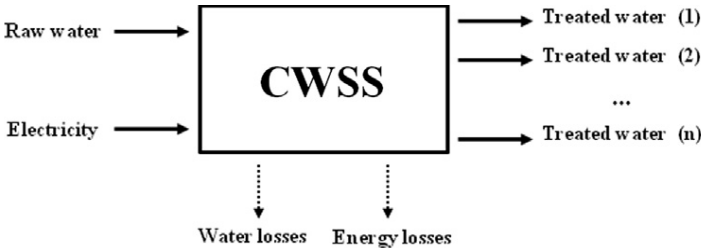

Figure 1: Energy and Hydraulic flows in a WSS (Vilanova, et al., 2014).

There are various methods to enhance energy efficiency in WSS. These methods are ranging from leakages control through simple monitoring operations to massive investments by reviewing the design of the infrastructures, equipment upgrade, pump system optimization, and real-time control. These Methods were classified into 3 major sub-categories in Figure 2 (Shankar, et al., 2016).

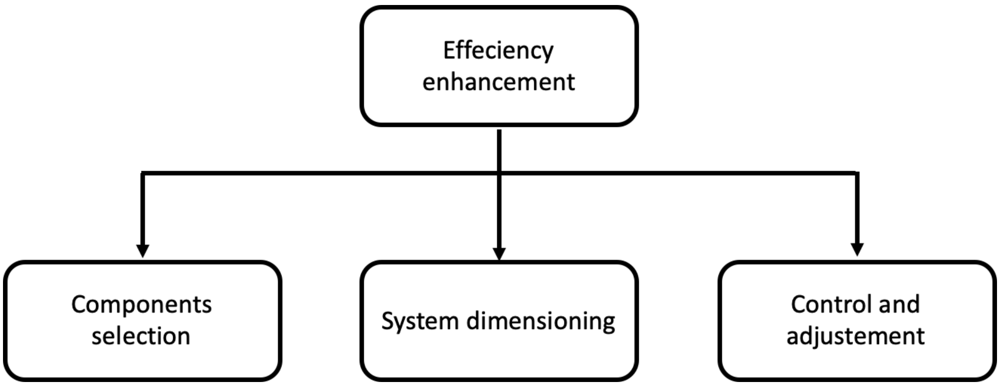

Figure 2: Efficiency enhancement opportunities in WSS.

2.1 Study area and description of the drinking water production system

The province of Taza is located in the center North of Morocco and it is one of the 9 provinces of the region of Fez-Meknes with a population of roughly 530 000 inhabitants (Morocco 2014). The water treatment plant of Tahla provides a large population of the province, mainly the urban areas. It is situated 60 km from the city of Taza. The plant treats the raw water of the Bab Louta reservoir.

The production system consists of a pumping station SP0 for raw water, a water treatment plant then another pumping station of SP3 to reach the city of Taza, which is the chief town of the province.

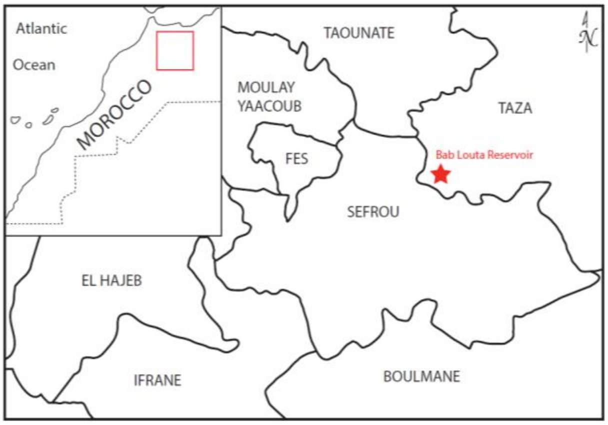

Figure 3: Location of the Bab Louta reservoir.

2.2 Description of the pumping station

The pumping station SP0 is located about 3km from the reservoir. It is responsible for overcoming the difference in altitude between the raw water intake and the water treatment plant. The water is taken by gravitation from the reservoir to the pumping station SP0 through a pipe of a nominal diameter (ND) of 700 mm which converge to a diameter of 500 mm just before reaching the SP0. Then, the water is pumped through a pipe of a nominal diameter of 600 mm to a tank destined to provide the water treatment plant with raw water, it is called RMC0 and having a capacity of 1000 m^3, which provides the water treatment plant with raw water through a pipe of a nominal diameter of 500mm (ONEE 2019).

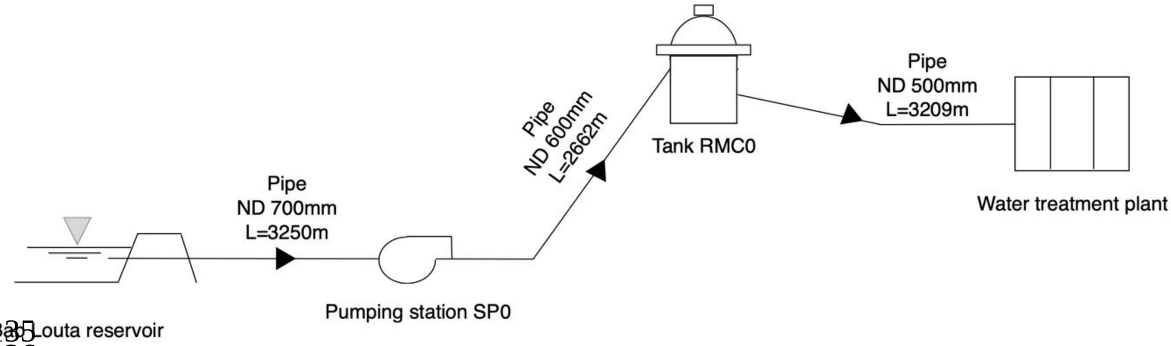

Figure 4: Location of the Bab Louta reservoir.

The pumping station, the object of the current study, is situated at an altitude of 498.20 m and tank RMC0 is situated at an altitude of 737.80m. SP0 consists of four pumps (Figure 5) with a flow of 457 m^3/h each and a manometric pressure of 209 m. They are manufactured by the company HIDROTECAR powered by an electric motor which is manufactured by LEROY SOMER with a nominal power of 455 kW (ONEE, 2019).

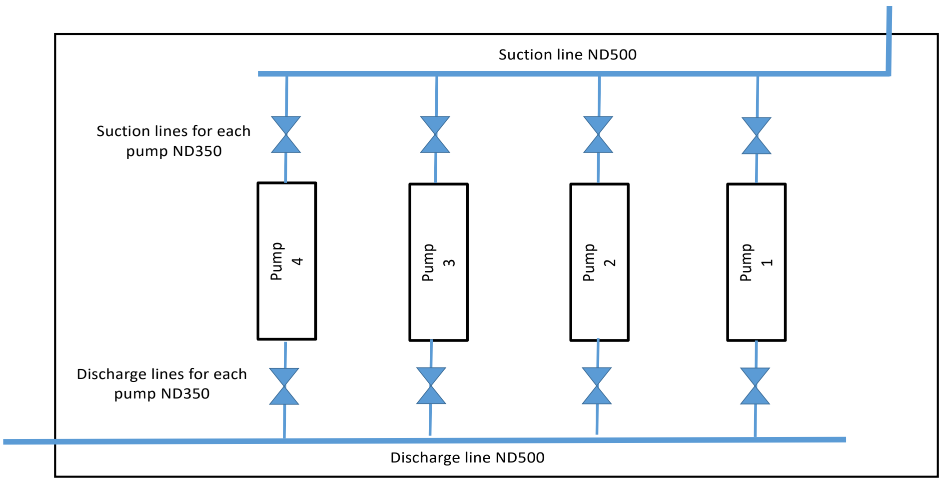

Figure 5: Hydromechanical scheme of the pumping station SP0.

2.3 Experimental procedure

To obtain a reliable model of our system, measurements of the key parameters influencing this system were conducted. These measurements were taken in the period between January 2015 and December 2018 daily for the parameters:

- P: the active energy consumed by the pumping station (measured by a wattmeter),
- Q: the reactive energy consumed by the pumping station (measured by a wattmeter),
- V: the daily produced volume (measured by an electromagnetic flowmeter),
- Cosφ: the power factor,
- HMG1: the pump operating time “1”,
- HMG2: the pump operating time “2”,
- HMG3: the pump operating time “3”,
- HMG4: the pump operating time “4”.

2.4 Modelling

This study aimed to use Multiple Linear Regression analysis to predict an output from a range of inputs. MLR model with multiple input variables can be expressed as follows (Longo et al., 2016):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n \qquad \text{Eq 1}$$

where :

- Y is the output variable;
- β_i are the regression parameters;
- X_i are the input variables.

In order to assess the influence of the included parameters on the cubic meters ratio produced by the pumping station, the following parameters were considered:

the active energy (EP), the reactive energy (EQ), the daily produced volume (V), the power factor (Cosφ), and the operating time of each pump (HMGi).

The effect of eight variables on the produced Kwh/m^3 ratio was evaluated (Table 1). Of note, 1388 experiments were conducted during 4 years.

Table 1: Problem characteristics

Objective of the study	The effect of the variables on the ratio KWh/m^3
Number of Variables	8
Number of experiments	1388
Number of the coefficients	8
Number of responses	1

Table 2: Measurements Summary

	E_P (kWh)	E_Q (kVar)	Cosφ	Production (m^3)	HMG1 (m)	HMG2 (m)	HMG3 (m)	HMG4 (m)	Ratio (kWh/m^3)
Mean	14 837.47	8 339.58	0.87	18 181.40	7.95	7.63	10.43	9.01	0.82
Standard deviation	9 158.63	8 529.85	0.05	5 081.13	8.90	5.88	11.51	8.85	0.43

In Table 2 there is a preview of descriptive statistics of the eight parameters, the mean and the standard deviation.

In the graphs below the distribution of the parameters is represented.

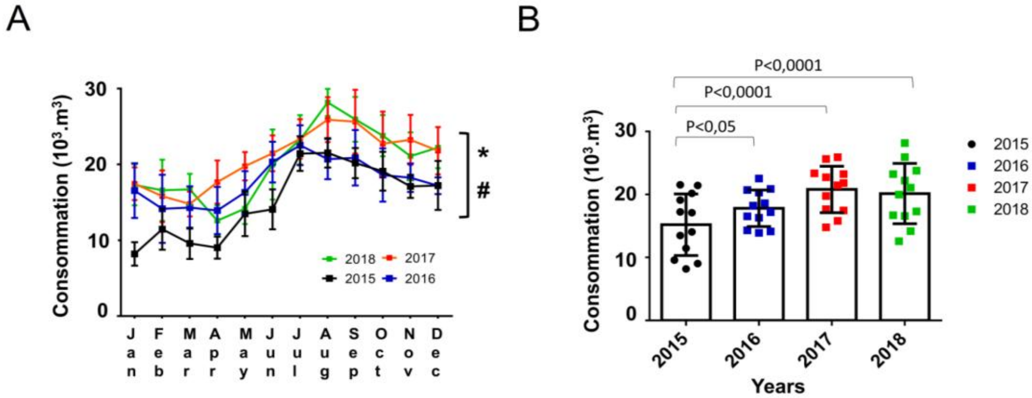

Figure 6: Representation the dataset collected of the variation of the consumptions through the months during period 2015 and 2018, (A) consumption variation through the months (*: p<0.05, consumption variation through the years; #: p<0.05, consumption variation through the months), (B) consumption variation through the years (P<0.05 is considered as statistically significant. Each point represents the consumption average of a given month during the year).

The dataset collected of the water production across the year for each year from 2015 to 2018 presented in figure 6-A, and it the trend shows that clearly demonstrates a higher production which reflects a higher consumption during the summer months in province of Taza. On the other hand, figure 6-B shows the evolution of the production through the years 2015 to 2018 and it has also the same increasing trend due to the continuous commissioning of new networks leading to a growing number of consumers.

3. **Results and discussion**

3.1 Statistical interpretation

From the correlation matrix (Table 3), it was found that the studied variables are strongly correlated:

- The active energy consumed by the pumping station was dependent on the production, reactive energy, pumps operating hours (4;1), power factor (Cosφ), HMG2, and HMG3 respectively.
- The reactive energy consumed is highly was dependents on the production, active energy, Cosφ, pumps operating hours (4;1;2;3) respectively.
- Cosφ was dependent on the reactive energy, the production, active energy, pumps operating hours (4;1;3;2) respectively.
- Production was dependent on the active energy, the reactive energy pumps operating hours (4;1;2;3).

Table 3: Correlation Matrice

Variable	Correlations (MATRICE CR without outliers) Marked correlations are significant at p < 0,05000 N=1388 (Case wise deletion of missing data)										
	Means	**Std. Dev.**	**P**	**Q**	**Cos phi**	**Prod**	**HMG1**	**HMG2**	**HMG3**	**HMG4**	**Ratio**
P	-0.022423	0.487229	1.000000	0.883071	-0.190305	0.939746	0.333340	0.202424	0.169873	0.625872	0.463030
Q	-0.027636	0.345041	0.883071	1.000000	-0.585775	0.883506	0.332525	0.170169	0.170505	0.612964	0.298339
Cosφ	0.026763	0.761882	-0.190305	-0.585775	1.000000	-0.316528	-0.197066	0.024943	-0.076562	-0.271961	0.193819
prod	0.000741	0.973866	0.939746	0.883506	-0.316528	1.000000	0.340309	0.222533	0.175107	0.613597	0.150594
HMG1	-0.001400	1.002171	0.333340	0.332525	-0.197066	0.340309	1.000000	-0.164919	-0.016355	0.236341	0.111733
HMG2	-0.001865	0.998758	0.202424	0.170169	0.024943	0.222533	-0.164919	1.000000	-0.086627	-0.141591	0.006575
HMG3	0.001362	1.003257	0.169873	0.170505	-0.076562	0.175107	-0.016355	-0.086627	1.000000	0.074751	0.036371
HMG4	-0.017745	0.708400	0.625872	0.612964	-0.271961	0.613597	0.236341	-0.141591	0.074751	1.000000	0.0257315
Ratio	-0.051987	0.197427	0.463030	0.298339	0.193819	0.150594	0.111733	0.006575	0.036371	0.0257315	1.000000

From the regression table (Table 4), it was concluded that the factors influencing the kWh/m^3 ratio in descending order are:

- Ratio is positively correlated with the active energy consumed by the pumps;
- Ratio is negatively correlated with the production;
- Ratio is positively correlated with Cosφ;
- Ratio is negatively correlated with the reactive energy consumed by the pumps;
- Ratio is positively correlated with the operating hours of pumps 1 and 4.

Table 4: Regression summary for dependent variable

N=1388	b*	Std. Err. of b*	B	Std. Err. of b	t (1379)
Intercept			-0.026073	0.001608	-16.2112
Prod	-2.52837	0.026164	-0.512563	0.005304	-96.6366
HMG_1	0.02116	0.008985	0.004168	0.001770	2.3546
HMG_2	0.01693	0.009327	0.003346	0.001844	1.8150
HMG_3	-0.00080	0.008299	-0.000157	0.001633	-0.0960
HMG_4	0.04287	0.011275	0.011948	0.003142	3.8023
P	2.85701	0.043662	1.157669	0.017692	65.4347
Q	-0.08315	0.040932	-0.047577	0.023421	-2.0314
Cosφ	0.09614	0.020445	-0.024913	0.005298	-4.7023

From the variance analysis (Table 5):

- Regression mean squares was superior to the residual mean squares;
- The Fischer value is very high;
- P-value is strictly inferior to 0.05.

These results validated our variance analysis.

Table 5: Analysis of Variance Table

Effect	Analysis of Variance; DV: Ratio				
	Sums of Squares	df	Mean Squares	F	p-value
Regress.	49.31283	8	6.164104	1790.044	0.00
Residual	4.74865	1379	0.003444		
Total	54.06149				

Multiple Linear Regression has shown that most of the experimental results are highly adjusted and the model is explanatory, considering the value of R^2 and the value standard error estimate which is very low (table 6).

Table 6: Summary statistics

Statistic	Value
R	0.95507172
R^2	0.91216198
Adjusted R^2	0.9116524
Fischer (8.1379)	1790
P	0
Standard error of estimate	0.0586817579

The final model is expressed by the equation below:

$$R_{ratio} = 2.85701\ P - 2.52837\ Prod + 0.09614\ Q - 0.08315\ Cos\varphi + 0.04287\ HMG1 + 0.02116\ HMG2. \qquad \text{Eq 2}$$

3.2 Technical Interpretations

From this analysis, this study concluded that in order to improve the ratio per produced cubic meter, a reduction of the active energy consumed by the pumps should be considered. As the function point of the pumps is already set by the system characteristics, only active energy could be reduced by the improvement of the pumps efficiency.

The ratio is negatively correlated with production means; i.e. there is an economy of scale. It means the most the production increases the least the ratio is.

The operating hours of the pumps 1 and 4 are positively correlated, which means that more these pumps are used higher more the ratio is higher. Therefore, it is recommended to use the pump 3, and if there is an operation of renovation of the pumping station, it is recommended to start with upgrading the pumps 1 and 4. In the event of a new investment, the company can increase the capacity of the RMC0 storage tank which, according to the model, will decrease the significantly the ration and also allows a load shifting to the off peak hours

The model which was elaborated in this study has a standard error of estimate of 0.05 and due to the lack of previous studies using multiple linear regression, we compared the results with a previous study involving five data-mining approaches (Kusiak et al., 2013). This study had for objective to model the energy consumption in a comparable case of a wastewater pumping station that has 3 pumps that transfer the wastewater to a treatment plant. Although there are differences between the flow capacities and the pressure with drinking water supply facilities but the approach remains the same. The five data mining approaches are the multi-layer, perceptron, neural network (MLP), the boosted-tree (regression) algorithm (BT), the random-forest algorithm (RF), the support-vector machine (SVM), and the k-nearest neighbor algorithm. These approaches had all provided more than 90% of accuracy which is the case in the model of this study. The benefit of our method goes beyond the control methods used in most of the optimization approaches which only provide a method to operate the system in an efficient way but don't account for other factors such as the aging of the pumps, factors that are crucial when upgrading the system.

4. Conclusions

A Linear multiple regression was conducted to assess and study the influence of multiple parameters on the energy consumption ratio per cubic meter water, involved in a water pumping station.

This unique approach has allowed evaluating the real response of the system relying on data that is measured over a 4 years period. Modelling the ratio will be a tool to take decisions on which pump should the work be done first. This method combined with a cash flow analysis, can help to take decisions on establishing priorities in case of renovations, to change the pumps 1 and 4 with more efficient pumps. The model would be useful when the next renovation will be undertaken by the office which will conduct a replacement of the pumps in the year 2024.

The above findings demonstrate the potential of method for solving real-time pump scheduling problems in large water distribution systems with many pumps. This, however, requires further work with other metaheuristic methods such as Genetic Algorithms before relevant conclusion can be made.

Data availability. The source data used for the illustrations of the cases are available upon request.

Conflict of interest. All authors declare no conflicts of interest in this paper.

References

Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B., Sliusarieva, A.: Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada WATER RESOURCES RESEARCH, VOL. 48, W01528, doi:10.1029/2010WR009945, 2012

Kusiak, A., Zeng, Y ., Zhang, Z.: Modeling and analysis of pumps in a wastewater treatment plant: A data-mining approach, Engineering Applications of Artificial Intelligence., 26, 7, 1643-1651, 2013.

Shankar, A., Umashankar, V. K., Paramasivam, S., Norbert., S. H.: A comprehensive review on energy efficiency enhancement initiatives in centrifugal pumping system, Applied Energy, Elsevier, 181, C, 495-513, 2016

Carravetta, A.; Giugni, M.; Malavasi, S.: Application of Innovative Technologies for Active Control and Energy Efficiency in Water Supply Systems. Water, 12, 3278. https://doi.org/10.3390/w12113278, 2020

Coelho, B. and Andrade-Campos, A.: Efficiency achievement in water supply systems-A review, Renewable and Sustainable Energy Reviews, 30, C, 59-84, 2014.

Moreira, D. F., Ramos, H. M.: Energy Cost Optimization in a Water Supply System Case Study, Journal of Energy, vol. 2013, Article ID 620698, 9 pages,. https://doi.org/10.1155/2013/620698, 2013

Durmus, K. E., Alptekin, Y., Suleyman Yigit, K., Kilic, F. C., Salih Eren, A., Cenk Celik.: Energy efficiency in pumps, Energy Conversion and Management, 49, 6, 1662-1673, 2008.

Fallside, F., Perry, P.F.: Hierarchical optimization of a water supply network, Proceedings of the Institution of Electrical Engineers, 122, 2, 202-208, 1975.

Giustolisi, O., Laucelli, D., Berardi, L.: Operational optimization: water losses versus energy costs, J. Hydraulic Eng. ASCE., 139, 4, 410-423, 2012.

Helena, M-J., Nargiz, S., Dragan, S.: Lost in optimisation of water distribution systems? A literature review of system operation, Environmental Modelling & Software, 93, 210-251, https://doi.org/10.1016/j.envsoft.2017.02.009, 2017

Morocco, 2014.: High Commission for Planning (Morocco). Morocco Population and Housing Census 2014, http://ghdx.healthdata.org/record/morocco-population-and-housing-census, 2014.

EIA, 2019.: International Energy Outlook 2019, https://www.eia.gov/outlooks/ieo, 2019.

Longo, S-A., Benedetto, M-B., Michael, C., Antonio,C., Andreas, F., Francesco, L., Juan, M. M-I., Miguel, S., Ana, H. A.: Monitoring and diagnosis of energy consumption in wastewater treatment plants. A state of the art and proposals for improvement, Applied Energy, Elsevier, 179, C, 1251-1268, 2016.

Luna, T., Ribau, J., Figueiredo, D., Alves. R: Improving energy efficiency in water supply systems with pump scheduling optimization, Journal of Cleaner Production 213, 342-356, 2019.

Nogueira, V., Mateus, R., Perrella, B., José, A.: Energy and hydraulic efficiency in conventional water supply systems, Renewable and Sustainable Energy Reviews, 30, C, 701-714, 2014.

ONEE-"National Office of Electricity and Drinking Water Morocco"-(2019) .: Exploitation report, 2019.

Ostfeld, A. and Tubaltzev, A.: Ant Colony Optimization for Least Cost Design and Operation of Pumping and Operation of Pumping Water Distribution Systems. Journal of Water Resources Planning and Management, 134, 107-118. http://dx.doi.org/10.1061/(ASCE)0733-9496(2008)134:2(107), 2008

Puleo, V., Morley, M., Freni, G., Savić, D., Multi-stage linear programming optimization for pump scheduling Procedia Engineering, 70, 1378-1385, 2014.

Plappally, A.K., Lienhard, J.H. V.: Energy requirements for water production, treatment, end use, reclamation, and disposal Renewable and Sustainable Energy Reviews 16, 4818–4848, 2012.

Rothausen, S., Conway, D. Greenhouse-gas emissions from energy use in the water sector. Nature Clim Change 1, 210–219, https://doi.org/10.1038/nclimate1147, 2011

Sarbu, I.: A Study of Energy Optimisation of Urban Water Distribution Systems Using Potential Elements. Water, 8, 593, 1-19. https://doi.org/10.3390/w8120593, 2016

Vilanova, N., Ricardo, M., Balestieri, P., Antônio, J.: Energy and hydraulic efficiency in conventional water supply systems, Renewable and Sustainable Energy Reviews, Elsevier, 30, C, 701-714, 2014.

Wu, P., Lai, Z., Wu, D., Wang, L.: Optimization research of parallel pump system for improving energy efficiency, Journal of Water Resources Planning and Management, 141, 8, 2015.

UN-Water 2020 World Water Development Report, https://www.unwater.org/publications/world-water-development-report-2020, 2020.

Zhou, Y., Lee, E. W. M., Wong, L. T., & Mui, K. W.: Environmental evaluation of pump replacement period in water supply systems of buildings. Journal of Building Engineering, 40, 102750, 2021.