REVIEW PAPER DWES-2021-16 (16 March 2022)

The paper illustrates a statistical analysis of operation of raw pumping station transporting the water from the source to the treatment plant, through a transmission main of approximately 3 km, based on the series of 4-year measurements of basic operational parameters.

GENERAL COMMENTS:

1. It is a case study-based paper without any specific scientific contribution.
2. I see no novelties claimed by the authors documented with any sufficient literature study.
3. The case could possibly be presented as a practitioner's paper but much is to be desired to bring it even to that level.
4. The background, the descriptions of the methodology, and the discussions and conclusions are pretty meagre. The whole structure of the paper is actually rather weak.
5. Although the text is not difficult to read, a further revision of English and explanations of used abbreviations is needed.
6. In this version, I cannot recommend the paper for publishing.

SPECIFIC COMMENTS (attached below the cut parts of the paper)

https://doi.org/10.5194/dwes-2021-16
Preprint. Discussion started: 20 January 2022

Open Access Drinking Water Engineering and Science Discussions

1 **Modeling of pump performance in a water pumping plant**
2
3 **Fouad LAAJINE[1,2], Mohammed MACHKOR[2], Driss MAZOUZI[1]**
4 [1]LRNE, multidisciplinary faculty of Taza, Sidi Mohamed Ben Abdellah University,
5 Fez, Morocco
6 [2]National Office of Electricity and Drinking Water Morocco
7 Correspondence: Driss MAZOUZI (driss.mazouzi@usmba.ac.ma)

SC 01: The title of the paper is not accurate description of the contents. I see no modelling component; it is a statistical analysis. Secondly, the term 'Water Pumping Plant' is confusing. I first thought that it was about clear water pumping station as an integral part of the water treatment plant, which is not the case. I would add the case study title to the revised paper title

14 of pumping systems. The main objective of this study is to produce a model which reflects
15 the real behaviour of a pumping system to help in taking decisions on which pump to use
16 First and which one to replace in case of a limited renovation. In order to do so, Multiple

SC02: I see no model in the study. It is a formula for statistical regression derived from the measured operational parameters. Not more, not less. I would certainly not understand how is that formula

used for definition of replacement strategies. What is meant with 'limited renovation'? All this is not explained in the paper. English spelling: 'First' with capital 'F'?

40 Pumps account for 80% to 90% of the energy consumption (Sarbu, 2016). By achieving
41 energy efficiency improvements measures, we can reduce the consumption by at least 25%
42 (Moreira, 2013). Very few studies were conducted before to simulate the real behaviour of
43 pumping systems and evaluate the influence of parameters such as the aging of the
44 components, which can induce a reduction of the pumps performance for up to 12% (Kaya,
45 2008).

SC03: It is awkward to generalize any percentages referred from the literature because these normally emerge from some cases i.e. under specific conditions, which are not elaborated here. The pump ageing is interesting aspect, but it is not defined in the paper. How do we measure/monitor it? Was this included in the objective?

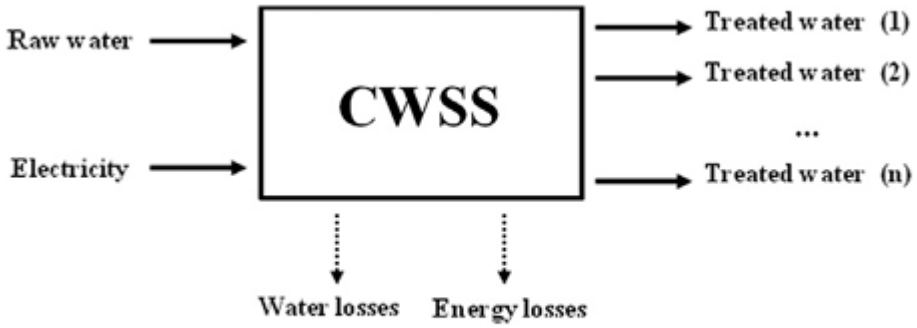

Figure 1: Energy and Hydraulic flows in a WSS

SC04: The drawing layout is confusing. It is mostly close to a water treatment plant. CWSS abbreviation does not stand because that one would also include transport and distribution infrastructure. On the other hand, the water and energy losses are indicated. Where they are originating from?

104 **1.4. Modelling:**
105 The aim of this study is to use Multiple linear regression, a wide popular technique to
106 predict an output from a range of inputs. MLP model with multiple input variables can be
107 expressed as following (Longo, 2016):

SC05: Why talking about the aim of the study in this place? What is the difference between the aim and objective? What is the exact meaning of MLP (should it be MLR?). English spelling: should be (' a widely popular technique'; 'Multiple linear regression' all words should start with capitals.

Table 1: Problem characteristics

Objective of the study	The effects
Number of Variables	8
Number of experiments	1388
Number of the coefficients	8
Number of responses	1

The table above summaries the objective of the study evaluating the effects of 8 variables on the response, which is the ratio of Kwh/m^3 produced. To get enough data, 1388 experiments were conducted during a period of 4 years.

SC06: The objective spelled in line 14 was to produce a model. Here it states that it is about 'the effects' (of what?). The table is confusing i.e. needs more elaboration: the difference between the variables and coefficients, what is meant with number of responses, etc.

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

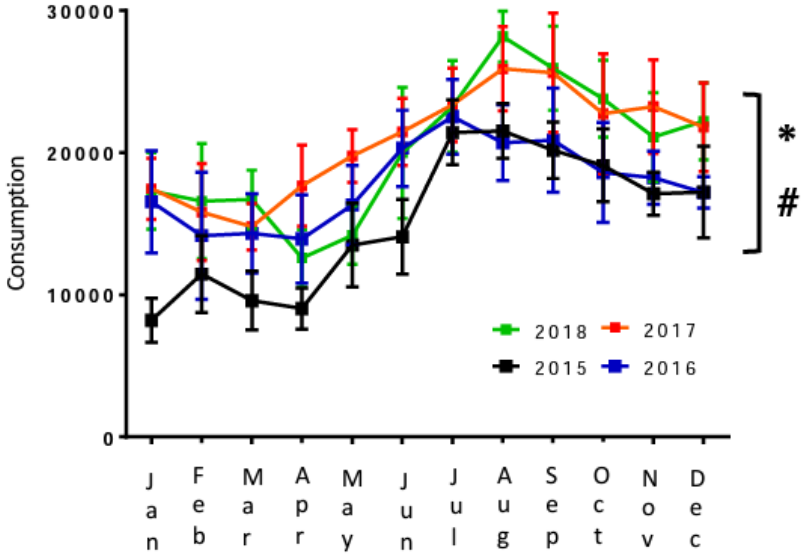

Figure 6: Representation of the variation of the consumptions through the years

SC07: Units are missing on Y-axis. Also, what is meant with 'Consumption'? Looking to the system layout in Fig. 4, it is more about a 'Production' in fact.

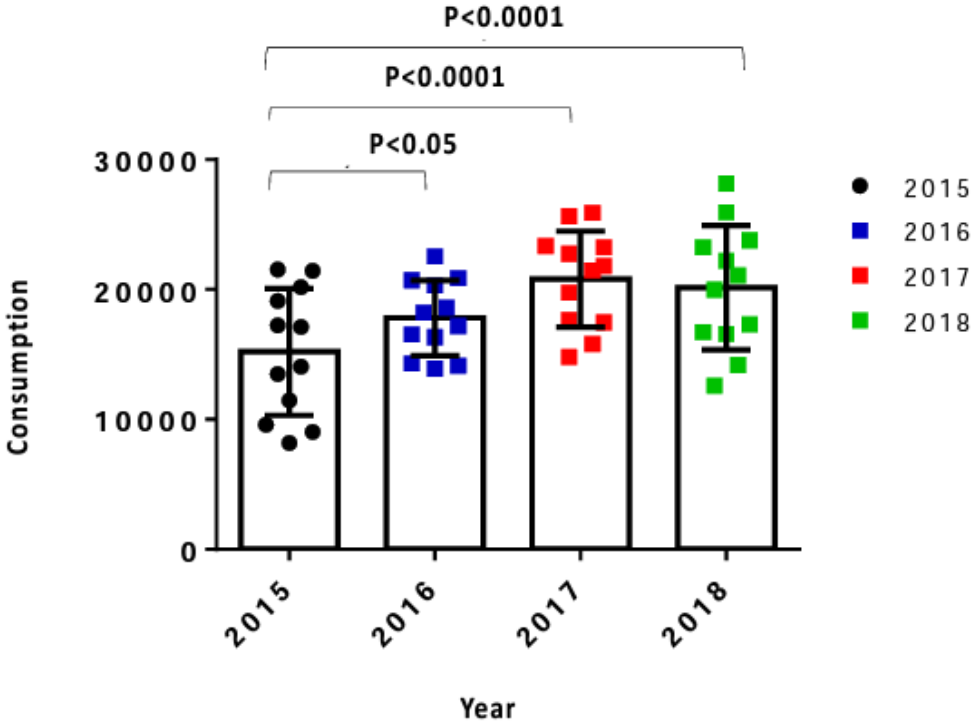

Figure 7: Box plots of the consumptions through the years

SC08: The same comment as SC07. Moreover, the meaning of P is not explained.

From the table of the regression summary (Table 4) it is conclude that the factors influencing the ratio in a 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 the CosPhi;
- Ratio is negatively correlated with the reactive energy consumed by the pumps;
- Ratio is positively correlated with the operating hours of the 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
E_p	2.85701	0.043662	1.157669	0.017692	65.4347
E_q	-0.08315	0.040932	-0.047577	0.023421	-2.0314
Cos Phi	0.09614	0.020445	-0.024913	0.005298	-4.7023

SC09: There is a repetitive mentioning of a ‘ratio’ but no explanation which one.

SC10: To which extent is the statistical analysis giving surprising or logical correlations? Could the relations be known even without doing it? The bullets only read the table, without real discussions.

181 The operating hours of the pumps 1 and 4 are positively correlated, which means that
182 the more we use them the higher the ratio gets, so we’d better use the other groups,
183 especially the pump 3, and if there is an operation of renovation of the pumping station, it is
184 recommended to start with changing the pumps 1 and 4.

SC11: The pumping station has four identical units. So obviously, shuffling their operation schedules does not interfere with the target hydraulic performance while it is ‘healthy’ for the lifetime of each pump. This is a common engineering logic. I do not understand what more we learn from the results in the tables in order to operate the pumps differently? The interpretation of the results is very superficial.

185 The model which is elaborated in this study has a standard error of estimate of 0.05 and
186 due to the lack of previous studies using multiple linear regression, we compared the results
187 with a study involving five data-mining approaches (Kusiak, 2013). The five data mining
188 approaches are the multi-layer, perceptron, neural network (MLP), the boosted-tree
189 (regression) algorithm (BT), the random-forest algorithm (RF), the support-vector machine
190 (SVM), and the k-nearest neighbour algorithm. These approaches had all provided more
191 than 90% of accuracy which is the case in the model of this study.

SC12: I see no evidence of any comparison in the paper. How can I trust?

196 This unique approach has allowed determining the real response of the system relying
197 on data that is measured over a 4 years period. Modelling the ratio will be a tool to take
198 decisions on which pump should the work be done first. This method combined with a cash
199 flow analysis, can help to take decisions on establishing priorities in case of renovations, to
200 change the pumps 1 and 4 with more efficient pumps.

SC13: What is ‘unique’? What is meant with ‘real response’? How do we really benefit from the measurements done to improve the operation of the pumps?

SC14: The suggested financial considerations should already be added to improve the substance of the paper.

SC15: I do still do not understand the rationale to replace ‘the pumps 1 and 4 with more efficient pumps’. Why they are currently worse than pumps 2 and 3 when they are all identical. Again, too superficial discussion of the results.