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.

Thank you for carefully reviewing our manuscript and providing a referee comment in the public review process. Your constructive comments will improve our manuscript. Please find our answers below in red

GENERAL COMMENTS:

1. It is a case study-based paper without any specific scientific contribution.

Response: Our specific scientific contribution in the paper is that, basically was fitted to model the produced energy consumption for pump efficiency of the pumping station and predict the next renovation of the pumps involved in a water pumping station. We believe the section in the aim of the introduction can be strengthened by highlighting these points raised by the reviewer:

"In this context, this work introduces a study based on statistical modelling by using multiple regression method to analyze the key factors affecting the efficiency of the pumping for drinking water production. Hence, this paper presents the results 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 energy consumption kWh/m^3 ratio according to the input key-parameters by using Real-Time-Data based on the collected data such as: the active energy and reactive energy consumed by the pumping station, the daily produced volume, the power factor and the pump operating time. Of other objectives, this study can also predict the next renovation of the pumps involved in a water pumping station."

2. I see no novelties claimed by the authors documented with any sufficient literature study.

Response: We will make the necessary revisions.

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.

Response: Thanks for your kind reminders. We revised, introduction, all the sentences, new reference... of the paper who you find in attached a new paper with all correction asked. We hope that the manuscript has been improved towards after this revision.

5. Although the text is not difficult to read, a further revision of English and explanations of used abbreviations is needed.

Response: We agree with the reviewer. We believe a reorganization and enhancement for next new paper after revision.

6. In this version, I cannot recommend the paper for publishing.

Response: These comments and suggestions will undoubtedly improve the impact and utility of our paper

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

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

Response: Thank you very much for your comment. We have changed the title to "Statistical Modelling Based on Multiple Linear Regression Analysis Method of Pumps Performance in the Pumping for Drinking Water Production"

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'?

Response: Thanks for your kind reminders. We revised the sentence as follows:

"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 by using data statistical analysis 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".

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?

Response: Thanks for your kind reminders. We agree with the reviewer's assessment. Accordingly, we revised the sentence as follows:

"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, the pump ageing, 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)."

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?

Response: We agree with the reviewer. We believe a reorganization and enhancement of this section can improve the text:

"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). Basically, the energy consumption in WSS is closely connected with water demand, generally, this consumption is associated to pumping systems and represents the largest share of energy consumption in the entire water sector. Therefore, it is interesting to develop energy and hydraulic model for to achieve high energy efficiency. To do this, the WSS systems should be evaluated a day-to-day analysis in terms of mass and energy (Figure 1)."

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.

Response: Thank you very much for the reminder. We have made revisions accordingly.

"Regression analysis, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. This study uses a Multiple Linear Regression (MLR) analysis to predict an output from a range of inputs. This means that the MLR analysis would make it possible to obtain a relationship between the key-parameters associated to the pumping for drinking water production and the produced kWh/m^3 ratio costs. MLR model with multiple input variables can be expressed as follows (Longo et al., 2016):"

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.

Response: Thank you very much for your previous comments that helped us improve this manuscript. We revised the sentence as follows:

"In order to assess the influence of the included parameters on the cubic meter ratio produced by the pumping station, the following key-parameters (input variables) were considered: the active energy (E_P), the reactive energy (E_Q), the daily produced volume (V), the power factor (Cosφ), and the operating time of each pump (HMG_i). Therefore, the principal parameter analysis is used to establish the evaluation model to achieve more objective and accurate analysis.

The effect of eight variables on the produced Kwh/m^3 ratio was evaluated. Of note, 1388 experiments were conducted during 4 years. The set of analysis data is summarized in Table 1."

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

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.

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

Response: Thank you for the nice reminder. We combined Figure 6 and 7 into one figure (Figure 6). We revised the sentence as follows

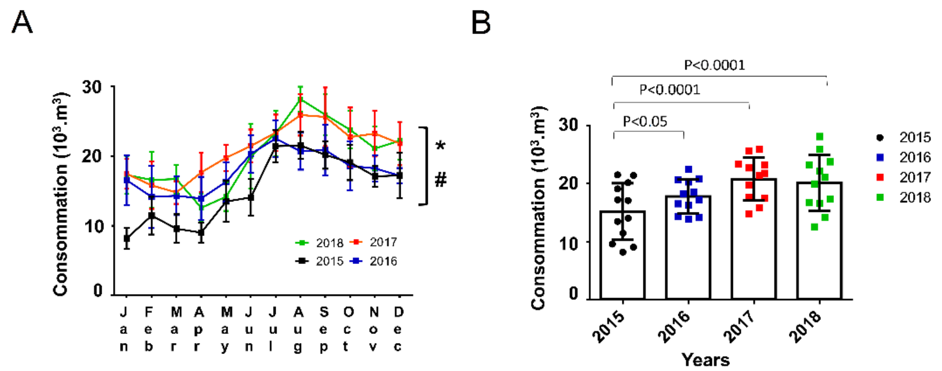

Figure 6: Representation the dataset collected of the variation of the consumptions of water versus the months during period 2015 and 2018, (A) consumption variation through the months (*: $P<0.05$, consumption variation through the years; #: $P<0.05$, consumption variation through the months), (B) consumption variation through the years (P-value statistical analysis; 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. The knowledge of such curves allows the consumptions of water to be assessed. In this way, the day, mouth and annual yield of the energy consumption related to pump station can be estimated. For analysis of this data, it was noticed that there is a higher production which reflects a higher consumption of water during the summer months in province of Taza (Figure 6-A). 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. The p-value of these statistical analysis is approximately null, showing that the explanation of the selected independent variables is statistically significant, considering a level of significance of 5%.

It can be seen from Table 1, 2 and Figure 6 that the original data has large differences and many influencing factors, and it is difficult to conduct comprehensive and systematic analysis by conventional methods. Therefore, the MLR analysis method is used to establish the evaluation model to achieve more objective and accurate analysis.

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.

Response: Thanks for your question, most of the update in this paper of this section:

"The correlation coefficients shown in a matrix (Table 3) are the results of statistical analyses for possible relationships between different parameters monitored. It was found that the studied variables are strongly correlated. Taking the above into the account in the energy consumption system of water supply, it is justified to include that:

- The active energy consumed by the pumping station was dependent on the production, reactive energy, pumps operating hours (4;1), power factor (Cosφ), HMG_2, and HMG_3 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)."

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.

Response: Thanks for the comments. More revisions have been made to the relevant parts in the:

“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. It allows comparing the results using Multiple Linear Regression models with pumping station to avoid excessive energy consumption. 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 suggested to start with upgrading the pumps 1 and 4, which may also depend on the aging of these pumps. 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.”

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

Response: Thanks for your question, the new section quoted above by the authors in the paper show the comparison the results

“The model which was elaborated in this study, was successfully validated in the statistical analysis. It shows that the R-square statistic reaching 0.91, and a standard error of estimate of 0.05. Thus, 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.

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.”

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.

Response: Thank you. It's a good question. The description is now revised to give better understanding:

"A Linear Multiple Regression was conducted to assess and study the influence of multiple parameters on the energy consumption ratio per cubic 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. To validate this model, we performed the performance test by determining the correlation R to show the link between the produced kWh/m^3 ratio and the following parameters such as active and reactive energies, the daily produced water volume, the power factor (Cosφ), and the operating time of each pump. the regression coefficients, thus validating the models. The final model describes accurately the consumption per cubic meter produced with a R-square statistic reaching 0.91.

After this study, we retain that the developed model can predict the energy consumption ratio per cubic meter water, involved in a water pumping station. Thus, 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, can more accurately and reasonably evaluate the efficiency pumping, according to the pumping unit model, motor power…

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

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

Response: We are grateful for this comment as it points to an important rationale of this study,

which concern the prediction to change the pumps. Typically, all the pumps are energy consumers. This is the reason why our final mathematical model includes the energy consumption of all pumps. However, it is of note that the energy consumption of these different pumps is not equivalent, particularly for pumps 1 and 4. For this reason, we have particularly underlined their energy consumption in the first version of this manuscript. In order to clarify this confusion in the revised manuscript, the discussion focusing only on 1 and 4 pumps was replaced by a discussion integrating the entire pumps.

The following paragraphs have been added to reflect this improved discussion:

“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. It allows comparing the results using Multiple Linear Regression models with pumping station to avoid excessive energy consumption, which is able to estimate the performance and to make a proper decision on the pumps. The results indicate 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 suggested to start with upgrading the pumps 1 and 4, which may also depend on the aging of these pumps. 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.”