International Journal of Chemical and Biochemical Sciences
(ISSN 2226-9614)

Journal Home page: www.iscientific.org/Journal.html

Wastewater Treatment Plant Redesign, and Modeling - A Case Study of Imintanoute WWTP's (Morocco)

A. CHAOUI [1], L. ELHAIMAR [1], M. LAAOUAN [2], R. BOURZIZA [1], K. SEBARI [1], W. ELKHOUMSI [1]*

[1] *IAV Hassan II, Rabat, Morocco.*
[2] *International Institute for Water and Sanitation, Rabat, Morocco.*

Abstract

Monitoring wastewater treatment plants (WWTPs) holds immense significance in safeguarding water quality and promoting environmental sustainability, especially in the face of escalating water scarcity. However, a significant challenge arises from missing data, creating a hurdle in establishing an optimal operational framework for WWTPs. This paper focuses on selecting suitable missing data imputation methods for wastewater quality parameters: Chemical Oxygen Demand (COD), Biochemical Oxygen Demand over 5 days (BOD5), and Total Suspended Solids (TSS). Linear trend imputation is found effective for TSS, while Expectation-Maximization (EM) is preferred for COD and BOD5. Imputation reveals pollutant loads surpassing the thresholds defined by the initial design criteria, prompting a reevaluation of Imintanoute WWTP's design and performance analysis. Overall, the paper propose an updated WWTP's configuration with pre-treatment and parallel secondary treatment lines, validated through modeling with SUMO22 software, that improves efficiencies by 5.59% for COD, 2.88% for TSS, and 7.37% for BOD5, while adhering to wastewater discharge standards.

Keywords: Wastewater Treatment Plant (WWTP), Missing data imputation methods, Modeling, SUMO.

Full length article **Corresponding Author*, e-mail: abdellahchaoui2000@gmail.com

1. Introduction

In response to urgent global imperatives regarding water scarcity, the advancement of wastewater treatment processes (WWTPs) has become indispensable to preserve water resources **[11]**. Monitoring WWTP performance play a pivotal role in addressing not only increasing safety requirements but also ensuring efficient operations and compliance with environmental and health regulations **[12,23].** Consequently, real-time information provides valuable insights for studying WWTP changes over time. This is crucial for predicting wastewater quality, evaluating the environment, and managing water resources **[2,10,16].** However, an ongoing challenge in real-time wastewater quality monitoring is the issue of missing data. This occurs due to various reasons, such as the absence of measurements, loss of recorded data, or considering available data as unsuitable **[19,23,26].** Technical problems like network issues, communication errors, and interruptions in data transmission also affect the completeness of data **[15]**.

Ignoring missing data, especially in the field of wastewater treatment, can lead to faulty conclusions, loss of precision and bias in analysis models **[12]**. Furthermore, it hampers the effectiveness of various modeling approaches that require complete information for all relevant variables **[25].**

In this context, imputation methods emerge as a potential solution to this issue. It involves filling in missing values with reasonable estimates **[19,20].** The field of imputation has been extensively studied and continues to be a focus of ongoing research. However, pinpointing the most effective imputation method for specific water quality variables proves to be a challenging task.

The primary contributions of this paper can be outlined as follows:

1. **Comparative Analysis of Missing Data Imputation Methods:** this paper conducts a comparative analysis of diverse techniques for imputing missing data. This analysis aims to identify the most suitable imputation method for specific water quality variables.
2. **Wastewater Plant redesign:** This redesign process based on the results of the imputation process, is guided by two established frameworks: the METCALF and EDDY guidelines **[17]**, as well as the ATV-DVWK-A281, 2001 standard **[7]**. This step in necessary for making enhancements to the plant's configuration to optimize its overall efficiency.
3. **Optimal Treatment Scenario Selection:** Via advanced modeling tools such as SUMO22, this paper performs an evaluation of various treatment scenarios. Through this

analysis, this contribution select the most optimal treatment scenario that attains the desired purification performance objectives effectively.

2. Materials and Methods

In this section, the paper build a systematic framework to outline the progression of critical steps of the WWTP analysis encompassing missing data imputation, WWTP redesign and WWTP modeling (Fig.1).

2.1. Study Site Description

The study site is a WWTP in Imintanoute City, Chichaoua Province, Morocco. The selected WWTP, designed in 2017, serves a population of 31,000 and treats up to 1,720 m3/day of municipal wastewater by 2025. The wastewater treatment process consisted of pre-treatment (screens), a grit chamber, anaerobic tanks and trickling filters. The biological treatment system was followed by a secondary clarifier before discharge as the final effluent. The influent designated for treatment emanates from the wastewater system serving the municipality of Imintanoute. Table 1 displays the statistical data encompassing the water quality of influent and effluent designated for the wastewater treatment plant in Imintanoute.

2.2. Water quality monitoring data

Monitoring the treatment processes of Imintanoute's WWTP for a significant period, from October 15, 2018 to October 15, 2019, has enabled us to collect a database containing several wastewater quality parameters (TSS, COD and BOD5). However, it is crucial to highlight that a significant proportion of the dataset is incomplete, with 71.9% of BOD5 values, 59% of TSS values, and 59% of COD records being missing. The missing data presents a considerable challenge in comprehensively assessing the operational efficacy of the plant.

2.3. Missing data imputation

Numerous imputation methods have been developed over the years to solve the problem of missing values but first it is essential to start by identifying the type of data missing.
In this paper case, IBM SPSS software is used to carry out this statistical imputation work. The IBM® SPSS® Missing Values module enables to manage missing data by simplifying the estimation of synthetic statistics and the application of imputation methods using advanced statistical algorithms.

2.3.1. Types of missing data

Missing data can be classified into three categories, according to Little & Rubin (1987) **[18]:**

- MCAR (missing completely at random): This type occurs when the locations of missing values in the data set are purely random. In other words, the absence of data from one variable has no relation to the missing values of that variable or the data from other variables, i.e., the probability of absence is the same for all variables. This probability therefore depends only on external parameters independent of that variable **[20].**
- MAR (Missing at random): This occurs when the location of missing values in the data set depends on other observed data. Thus, data are not missing completely at random; if the probability of absence is related to one or more other observed variables **[20].**
- MNAR (Missing not at random): This type of missing data occurs when the location of the missing values in the data set depends on the missing values themselves, i.e. the probability of absence depends on the variable in question **[19].**

2.3.2. Imputation methods

- ***Mean Imputation***

Mean imputation consists in replacing each missing value of a variable by the mean value of the set of responses obtained for the same variable.

- ***Median imputation of neighboring points***

The median imputation method for neighboring points replaces missing values with the median of the valid values surrounding them. The neighboring point interval is the number of valid values above and below the missing value used to calculate the median

- ***Linear interpolation imputation***

This imputation method involves replacing missing values by linear interpolation. The last valid value before the missing value and the first valid value after the missing value are used for interpolation. If the first or last observation in the series contains a missing value, it is not replaced.

- ***Imputation by linear point trend***

The method of replacing missing values by the linear trend at the point consists in estimating missing values based on the linear relationship existing at this specific point. In this way, it seeks to exploit the linear trend observed at the point concerned to make an appropriate estimate.

- ***Imputation by Expectation-Maximization (EM):***

This method assumes a distribution for partially missing data and uses this distribution to make inferences. Each iteration of the process comprises an E step (Expectation) and an M step (Maximization).
Step E aims to conditionally estimate "missing" values based on observed values and current parameter estimates. These conditional estimates are then substituted for the "missing" values. In step M, the maximum likelihood estimates of the parameters are calculated on the assumption that the missing data have been filled.

- ***Linear regression imputation***

This method implements several linear regression estimates, and can improve estimates by using random components. For each predicted value, the procedure has the option of adding a residual from a randomly selected complete observation, a random deviation according to a normal distribution, or a random deviation (adjusted according to the square root of the mean of the squares of the residuals) from a t-shaped distribution. This approach makes it possible to take account of uncertainty and variability in estimates, by integrating random elements into the prediction process.

2.3.3. Evaluation Metrics for Imputation Methods

The performance of the recovering missing data is assessed based on the mean Absolute Error (MAE), the Mean

Squared Error (MSE) and the Root Mean Squared Error (RMSE).

$$\mathbf{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \qquad (1)$$

$$\mathbf{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \qquad (2)$$

$$\mathbf{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2} \qquad (3)$$

Where $\boldsymbol{y_i}$ and $\boldsymbol{\hat{y}_i}$ are the original and the predicted values of the water quality variable y and N is the total number of observations.

2.3. WWTPs redesign

The redesign of the Imintanoute's WWTP adheres to the guidelines specified in the METCALF and EDDY reference **[17]**, along with the ATV-DVWK-A281, 2001 standard **[7]**. These well-established standards provide a methodological framework essential for adapting treatment plants to specific requirements and optimizing their purification performance.

2.4. WWTPs Modeling

2.4.1. Modeling Approach

Sumo22, developed by the company Dynamita, is a robust, open-source, and versatile simulation software designed for environmental modeling, with a particular focus on the modeling of municipal and industrial wastewater treatment plants **[9]**.

To effectively model the operation of a wastewater treatment plant in SUMO22, a systematic approach involves the following series of methodical steps **[9]**:

- Configuration: This initial phase entails defining the structural layout of the treatment plant that will be represented within the model. This involves setting up the various components, such as pre-treatment structures, bioreactors, digesters etc., as well as the pipes connecting them.
- Model: This step is optional and involves selecting the mathematical model to describe the biological, chemical and physical processes. If no parameters are changed, the default parameters will be used - the model will be functional.
- Tools: users can select calculation parameters such as sludge retention time, proportional flow dependency, yields and others.
- Inputs: This aspect involves inputting the plant's data, which can either be constant or dynamic.
- Outputs: In this stage, the desired presentation format for the simulation results is defined. This can include tables, diagrams, charts, and other formats.
- Simulation: This phase represents the conclusion of the modeling procedure, where results are presented according to the format selected in the previous step. It provides valuable insights into the functioning and the performance of the wastewater treatment plant as represented by the SUMO22 model.

2.4.2. Model calibration and validation

Since modeling is an approximation of reality, it is expected that some values will differ from field observations. The aim of calibration is not a perfect match between simulation results and empirical data, but rather to reduce the error resulting from the differences between these two sets of values **[10]**. In this paper, the key adjustment for model calibration revolves around the effluent fractionation, integrated by default into SUMO22 via the "Influent tool"**[7].** The model chosen is based on chemical oxygen demand (COD), as advices by experts from Dynamita, given its role in measuring pollution levels.

Fractional adjustment focuses on COD and identifies several COD fractions, each affecting the values of Biochemical Oxygen Demand (BOD5) and Total Suspended Solids (TSS). To ensure the precision of the calibration process and validate the model, a permissible margin of error has been established, set at ±5 mg/l for TSS and ±3 mg/l for COD **[22].**

3. Results and Discussion

3.1. WWTP data imputation results

3.1.1. Type of missing data

To identify the type of the WWTP missing data, the chi-square statistic ($\chi 2$) is used for Little's MCAR test [18] to assess whether the WWTP data (COD, TSS and BOD5) is missing completely at random (MCAR). The null hypothesis is that the data is MCAR. Rejection of the null hypothesis provides sufficient evidence to indicate that the data are not MCAR. The IBM® SPSS® Missing Values module is utilized to conduct the chi-square statistic ($\chi 2$) analysis for Little's MCAR test, yielding the resulting outcomes table 2. Table 2 showed that the data from the station subject to imputation are MCAR since the p-value is equal to 0.319, significantly exceeding the 0.05 threshold.

3.1.2. Imputation methods results

The imputation process, using IBM® SPSS® Missing Values generates substantial outcomes, given that it involves replacing each absent value with an estimated value, resulting in a total of 366 values for each variable. Consequently, for every method employed, the paper display the median value derived from all imputation results in the table 3:

3.1.3. Comparative analysis of imputation method's accuracy

Tables 4,5 and 6 assesses the performance of six distinct imputation methods in estimating missing values of each water quality variable (TSS, COD and BOD5), using Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE) metrics to quantify the accuracy of the imputation. Notably, the comparative analysis of the six imputation methods demonstrates that while the linear point trend imputation is identified as the preferred technique for imputing missing values within the TSS variable, the Expectation-Maximization (EM) imputation method is particularly noteworthy for its efficacy in addressing missing values pertaining to the COD and BOD5 variables. Table 7 illustrates that the pollutant load values for

TSS, COD, and BOD5, resulting from the imputation process of the dataset 2018/2019, surpass the dimensions originally employed for plant design. This difference underlines the necessity to undertake plant redesign in order to ensure the attainment of desired purification performance levels. It is important to note that the redesign flow rate is calculated as the median value of the observed values within the same year.

3.2. WWTP redesign

The principal objective of this redesign initiative was to evaluate the treatment capacity of the plant regarding the loads of total suspended solids (TSS), chemical oxygen demand (COD), and 5-day biochemical oxygen demand (BOD5). Remarkably, these loads were found to exceed the benchmarks considered during the preceding design endeavor. In Table 8, the redesign of the Imintanoute wastewater treatment plant, predicated upon operational data spanning the 2018/2019 period, has yielded diminished dimensions across all structural components as compared to the original design. When comparing the dimensions of the plant's structures, it is imperative to note that an equal number of units has been retained across the various plant components for the purpose of this analysis. The disparity between the old and new configurations of the WWTP components can be attributed to the reduction in the influent flow rate, from 1,720 m^3/d for the old design to 1,246 m^3/d for the new design, representing a divergence of 27.56%. This variation, notably close to the difference observed for the various WWTP’s components, except the anaerobic ponds, where the said difference reaches 59%, embodies a more marked dissimilarity for this specific component. To address this challenge, a pertinent alternative strategy is proposed, entailing the retention of the existing plant configuration while effecting targeted adjustments. This proposed solution involves that one of the three anaerobic tanks will be temporarily deactivated. Consequently, the new configuration integrates pretreatment measures, succeeded by the establishment of two distinct treatment lines. Each of these treatment lines comprises an anaerobic tank, a bacterial bed, and a clarifier. The efficacy of this alternative approach remains assured until the flow attains the projected threshold of 2016 m^3/d by the year 2035, along with the anticipated Biological Oxygen Demand (BOD5) load of 600 mg/l (Table 09). Notably, this design modification would yield considerable reductions in the expenses associated with the periodic maintenance and sludge clearance from the decommissioned anaerobic tank, a requisite procedure occurring annually as required. The next step is to model the various WWTP configurations, using a calibrated model to guarantee maximum optimization of the WWTP's performance.

3.3. WWTP modeling

Figure 2 shows the original configuration of the Imintanoute’s wastewater treatment plant in the SUMO22 modeling software interface. The paper engaged in the modeling of the original station's configuration utilizing two distinct sets of data. The initial dataset pertains to the original design data, while the second dataset comprises the new data derived from the imputation procedure conducted on the data for the year 2018/2019. This approach allows us to comprehensively analyze and compare the original WWTP performance under both the original and updated datasets.

The new wastewater treatment plant in question is the configuration recommended in the previous section, which takes the form of pre-treatment followed by the presence of two parallel operational chains, each consisting of an anaerobic tank, a bacterial bed and finally a clarifier. Figure 3 shows the new suggested configuration of the wastewater treatment plant in the sumo22 modeling software interface.The modeling of this new configuration is executed in two phases. Firstly, it employs the new dataset for the year 2018/2019. Subsequently, data from table 09, represent the new WWTP limits, are employed to assess its efficacy in treating wastewater influents up to the year 2035. The results of the various scenarios show an appreciable degree of compliance with the requirements for discharge of treated water into the natural environment. So, in order to distinguish between each treated case and evaluate their performance, it becomes imperative to undertake a comparative analysis taking into consideration the concentration levels of polluting agents as well as their purification yields at the plant outlet.

The effluent load concentrations (COD, BOD5 and TSS) for the four scenarios examined are illustrated in figure 4. For the old WWTP, the residual concentrations of pollutant loads following assimilation of the new data turn out to be of lesser magnitude than those observed when using the old data. This disparity can be attributed to the fact that, even with an increase in pollutant loads, the influent flow rate to the plant is lower than its previous level, generating no substantial variation. As a result, it can be concluded that the wastewater treatment plant (WWTP) demonstrates its ability to treat the wastewater received, even in the context of this increase in pollutant loads observed during the 2018/2019 operating year.

the new WWTP configuration, employing the new dataset reveals a considerable reduction in pollutant load concentrations at the outlet, as opposed to the previous models. By utilizing the projected threshold values for 2035 (table 09), the efficacy of the new configuration in managing these substantial pollutant loads becomes evident. However, there has been an increase in effluent load concentrations, attributed to the notable escalation in pollutant loads. Figure 5 shows a comparison of the performance of the different variants of the imintanooute plant. The new configuration of the Imintanoute WWTP exhibits heightened treatment efficiencies in contrast to the treatment outcomes previously documented with the plant's original design. Specifically, these enhancements encompass a 5.59% refinement in COD reduction, a 2.88% enhancement in TSS reduction, and a 7.37% improvement in BOD5 reduction. These advancements underscore the pivotal significance of continuously monitoring wastewater characteristics throughout plant operation, thereby facilitating the optimization of plant functioning and the augmentation of treatment efficacy.

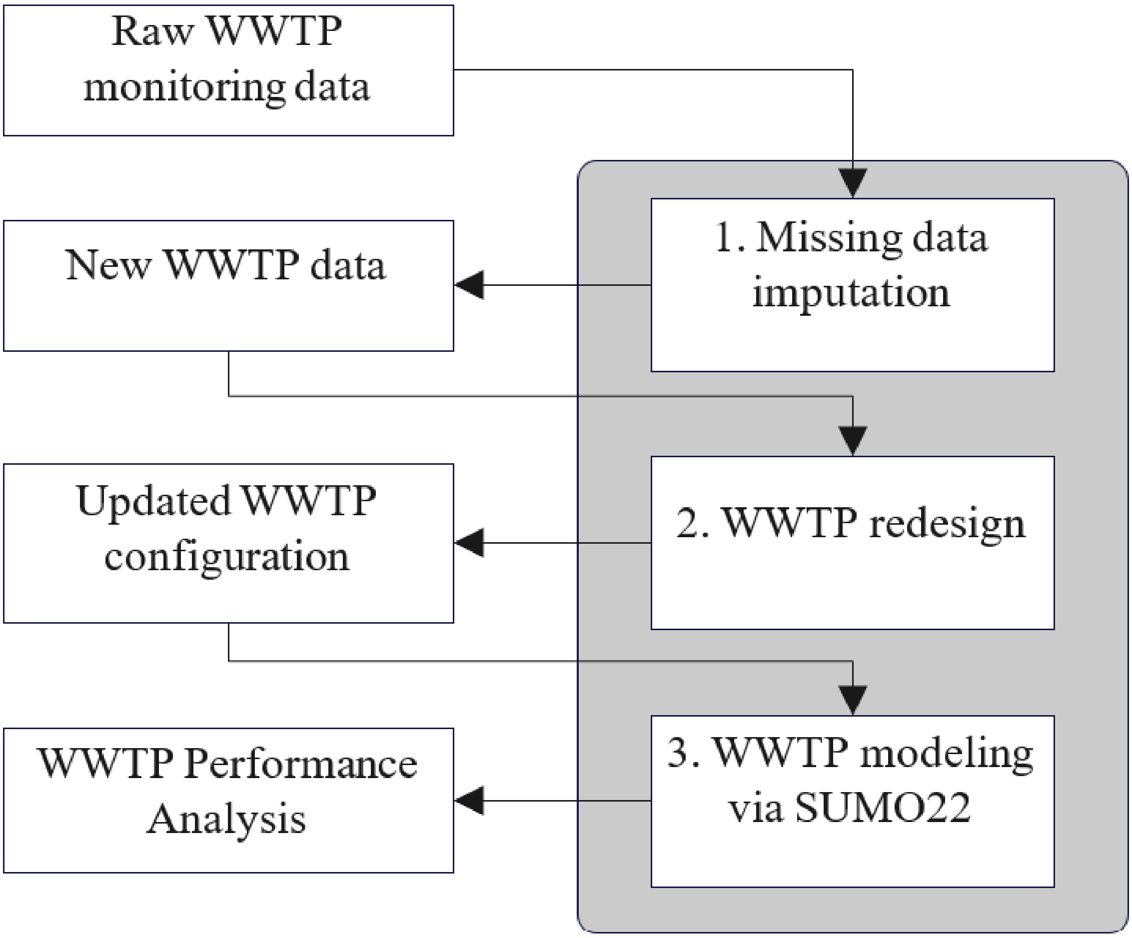

Figure 1: Systematic framework for WWTP analysis

Table 1: The influent and effluent designated for the WWTP in Imintanoute

Parameter		Units	designated value
Influent	**flow rate**	m3/j	1720
	BOD5	mg/l	523
	COD	mg/l	1046
	TSS	mg/l	639
Effluent	**BOD5**	mg/l	50
	COD	mg/l	150
	TSS	mg/l	75

Table 2: Chi-square statistic for MCAR ($\chi2$) results

Chi-square	2,283
DDL	2
P-value	0,319

Table 3: Median Values of Imputed Data across Multiple Imputation Methods

Wastewater parameters	TSS	COD	BOD5
Mean Imputation	642,56	1151,53	603,45
Median imputation of neighboring points	620	1172	570
Linear interpolation imputation	644,515	1173,5	565
Imputation by linear point trend	632,26	1147,64	602,625
EM imputation	642,56	1151,53	590,06
Linear regression imputation	636	1142,195	604,145

Table 4: Imputation accuracy for TSS

	TSS		
	MAE	**MSE**	**RMSE**
Mean Imputation	39,34	6613,45	81,32
Median imputation of neighboring points	43,68	12405,94	111,38
Linear interpolation imputation	50,24	12057,29	109,81
Imputation by linear point trend	34,28	5585,69	74,74
EM imputation	39,34	6613,45	81,32
Linear regression imputation	86,63	33855,62	184,00

Table 5: Imputation accuracy for COD

	COD		
	MAE	**MSE**	**RMSE**
Mean Imputation	61,41	23643,98	153,77
Median imputation of neighboring points	66,65	29964,28	173,10
Linear interpolation imputation	68,41	25042,99	158,25
Imputation by linear point trend	65,23	26202,24	161,87
EM imputation	61,41	23643,98	153,77
Linear regression imputation	110,53	60839,38	246,66

Table 6: Imputation accuracy for BOD5

	BOD5		
	MAE	**MSE**	**RMSE**
Mean Imputation	37,76	8333,00	91,29
Median imputation of neighboring points	59,26	20735,54	144,00
Linear interpolation imputation	53,97	18032,72	134,29
Imputation by linear point trend	37,54	8288,17	91,04
EM imputation	33,76	5674,63	75,33
Linear regression imputation	57,47	19428,29	139,39

Table 7: The new values of Flow rate, TSS, COD and BOD5

Variable	Flow rate (m3/j)	TSS (mg/l)	COD (mg/l)	BDO5 (mg/l)
New Value	1246	644,515	1151,53	590,06
Old values	1720	639	1046	523
%	-27,56	0,86	10,09	12,82

Table 8: Results of redesigning the components of the WWTP of Imintanoute

WWTP components	Parameter	Unit	New design	Old design	Difference
Coarse Screen	**grid surface**	m2	1,74	2,41	28%
	Grid refusal	l/j	125,34	142,01	12%
Fine Screen	**grid surface**	m2	0,44	0,61	28%
	Grid refusal	l/j	752,05	852,06	12%
Grit Chamber	**Volume**	m3	27	33,75	20%
	Grit Production	kg/j	307,31	461,44	33%
Anaerobic Pond	**Volume**	m3	1897,5	4579,2	59%
Trickling filter	**Volume**	m3	425,29	509,69	17%
Clarifiers	**Surface**	m2	95	143,14	34%

Table 9: The new configuration limits

Inhabitant equivalent	Flow rate (m3/j)	BOD5 (mg/l)	COD (mg/l)	TSS (mg/l)
40 000	2016	600	1253	671

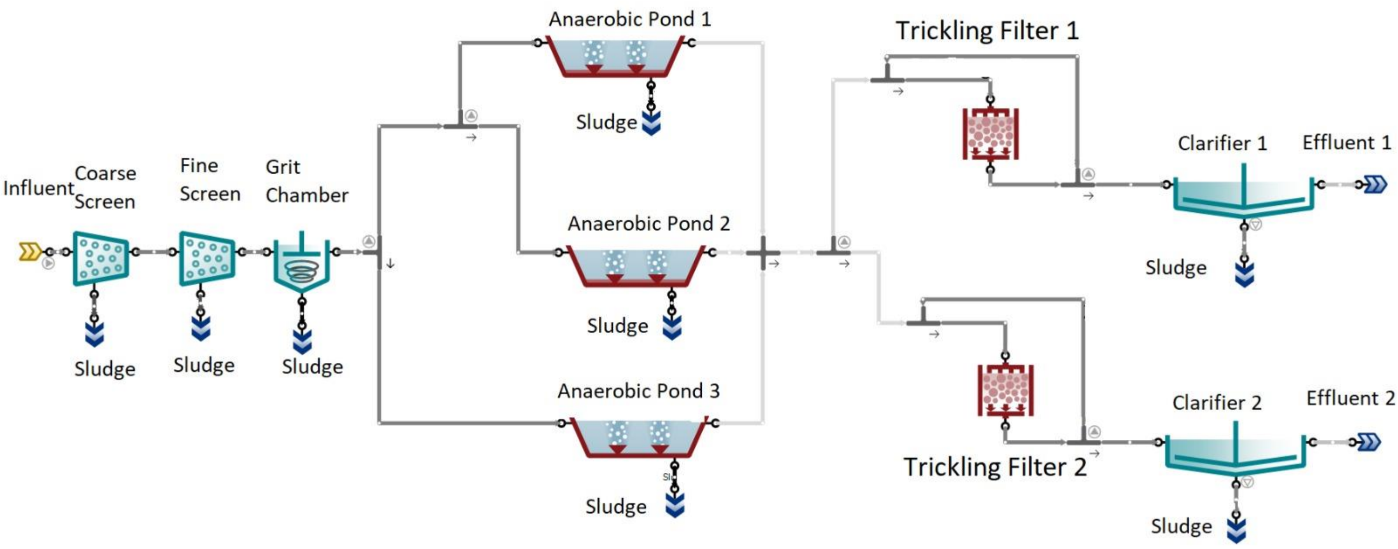

Figure 2 : Original Imintanoute WWTP

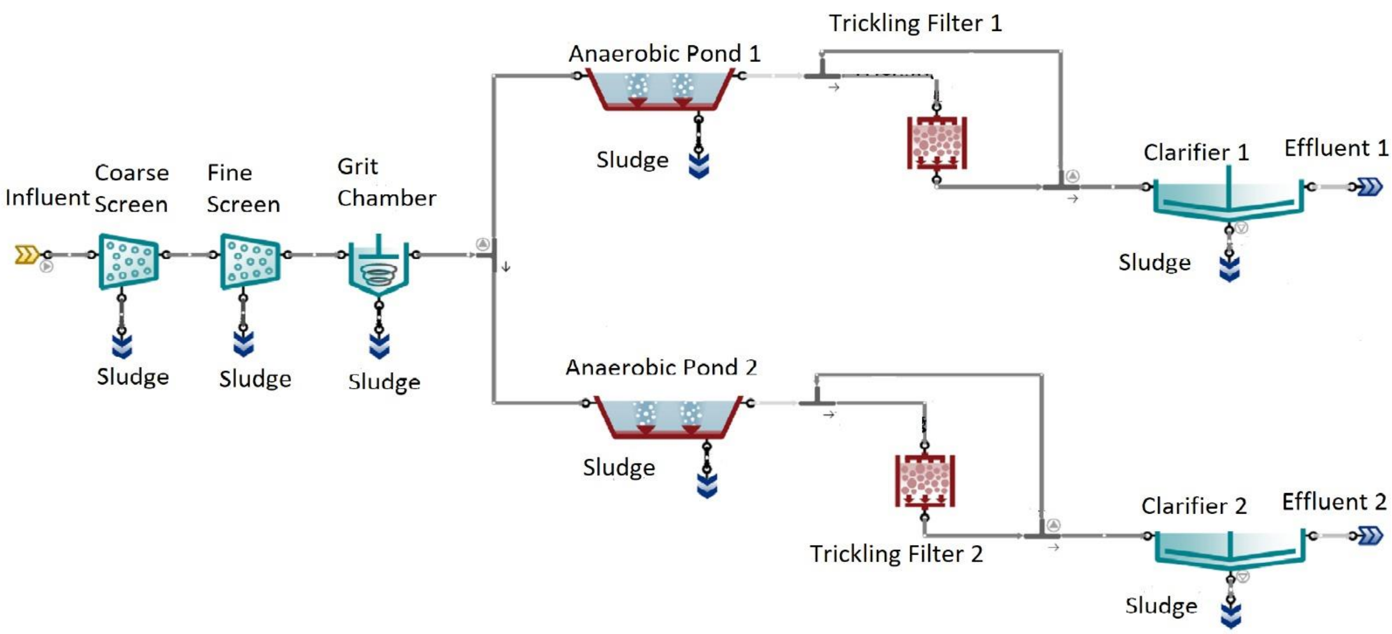

Figure 3 : New Imintanoute WWTP configuration

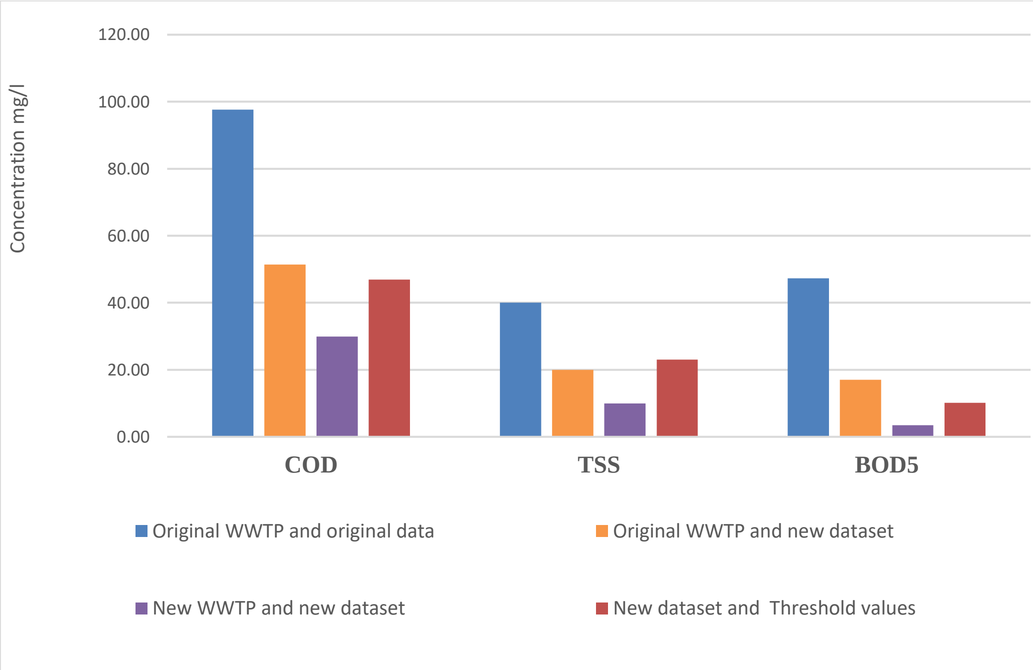

Figure 4 : The effluent load concentrations (COD, BOD5 and TSS) for the four cases examined

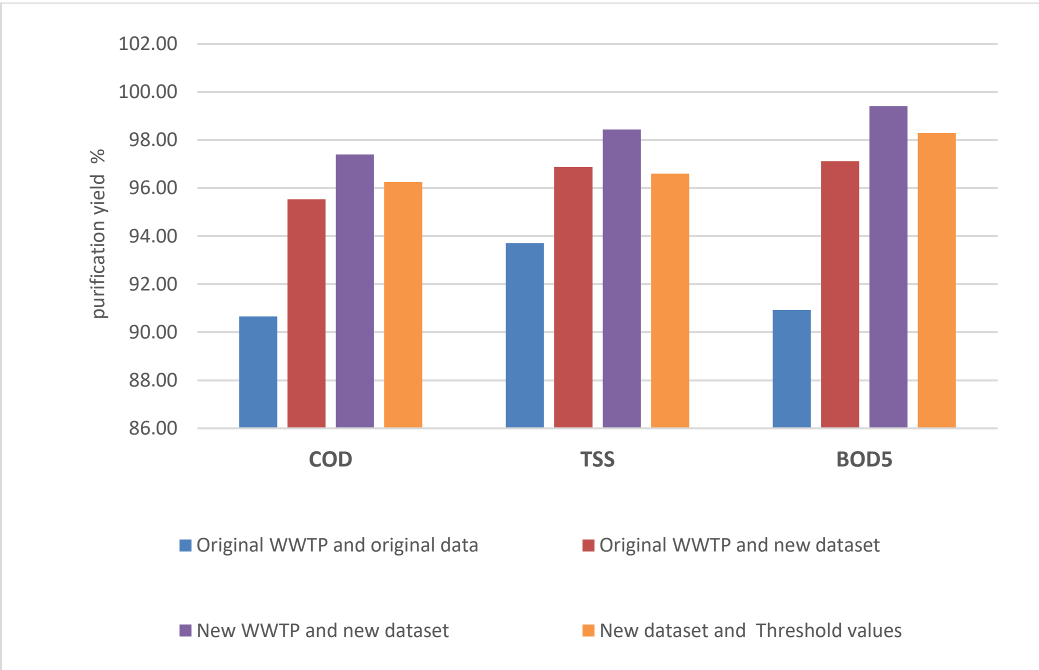

Figure 5 : The performance of the different variants of the Imintanooute plant.

4. Conclusion

Addressing the issue of missing data within wastewater treatment plant databases is integral to effective wastewater management. This paper focus on handling the missing data in the Imintanoute wastewater treatment plant's 2018/2019 database. This data gap, stemming from diverse factors like budget constraints and technical limitations, disrupts information reliability. The paper primary objective is to select appropriate imputation methods for key parameters: Chemical Oxygen Demand (COD), 5-day Biochemical Oxygen Demand (BOD5), and Total Suspended Solids (TSS). The paper findings, validated for accuracy, highlight that linear trend imputation is the optimal method for handling missing TSS values. On the other hand, Estimation-Maximization (EM) proves to be more suitable when dealing with COD and BOD5 parameters. Imputed data reveals pollutant loads exceeding initial Imintanoute WWTP's design criteria for TSS, COD, and BOD5, prompting a reconsideration of plant design and purification performance. Therefore, a redesign process is undertaken and lead to a reconfiguration proposal, improving operational efficiency, reducing costs and guaranteeing better effluent quality. Modeling with Dynamita's SUMO22 software confirms the viability of the optimized configuration. The results obtained from the simulations demonstrate improved purification yields and also ensure compliance with wastewater discharge standards. Moreover, this contribution modeling indicates that this new WWTP configuration holds even more promising treatment efficiencies and operational stability, under extreme conditions. Adopting this optimized configuration offers potential cost savings in maintenance, and anaerobic tank management.

References

[1] J.P. Boltz, E. Morgenroth, & D. Sen. (2010). Mathematical modelling of biofilms and biofilm reactors for engineering design. Water Science and Technology, 62(8), 1821-1836. https://doi.org/10.2166/wst.2010.076

[2] J.-P. Canler (2009). Aide au diagnostic des stations d'épuration par l'observation microscopique des boues activées. Cemagref.

[3] F. Cui, M. Kim, W. Lee, C. Park, & M. Kim (2021). Pseudo-analytical solutions for multi-species biofilm model of aerobic granular sludge. Environmental Technology, 42(22), 3421-3431. https://doi.org/10.1080/09593330.2020.1733673

[4] E. Curnow, J. R. Carpenter, J. E. Heron, R. P. Cornish, S. Rach, V. Didelez, M. Langeheine, & K. Tilling (2022). Multiple imputation of missing data under missing at random : Compatible imputation models are not sufficient to avoid bias [Preprint]. Epidemiology. https://doi.org/10.1101/2022.11.04.22281883

[5] C. David & K. G. (2014). Mathematical Modeling in the USMA Curriculum.

[6] H. Dhaoudi (2008). Traitement des Eaux Usées Urbaines : Les procédés biologiques d'épuration.

[7] Dimensioning of trickling filters and rotating biological contactors. (2001). ATV-DVWK, Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall e.V.

[8] Donders,A. R. T. Donders, G. J. M. G. van der Heijden, T. Stijnen, & K. G. M. Moons (2006). « Review: A gentle introduction to imputation of missing values », Journal of Clinical Epidemiology, vol. 59, no 10, p. 1087-1091.

[9] Dynamita (2022) Sumo User Manual.

[10] K. Elmaadawy, M. A. Elaziz, A. H. Elsheikh, A. Moawad, B. Liu, & S. Lu (2021). Utilization of random vector functional link integrated with manta ray foraging optimization for effluent prediction of wastewater treatment plant. Journal of Environmental Management, 298, 113520. https://doi.org/10.1016/j.jenvman.2021.113520

[11] H. Han, Z. Liu, Y. Hou, & J. Qiao (2019). Data-driven multiobjective predictive control for wastewater treatment process. IEEE Transactions on Industrial Informatics, 16(4), 2767-2775.

[12] H. Han, M. Sun, H. Han, X. Wu, & J. Qiao (2023). Univariate imputation method for recovering missing data in wastewater treatment process. Chinese Journal of Chemical Engineering, 53, 201-210.

[13] H. Horn & S. Lackner (2014). Modeling of Biofilm Systems : A Review. In K. Muffler & R. Ulber (Éds.), Productive Biofilms (Vol. 146, p. 53-76). Springer International Publishing. https://doi.org/10.1007/10_2014_275

[14] T. Hug, L. Benedetti, E. R. Hall, B. R. Johnson, E. F. Morgenroth, I. NOPENS, L. Rieger, A. R. Shaw, & P. A. Vanrolleghem (2009). Mathematical models in teaching and training : mismatch between education and requirements for jobs. Water Science and Technology, 59(4), 745–753.

[15] M. Jabir (2018). Comparaison de méthodes d'imputation des données manquantes appliquées à la base nationale sur les collisions (Doctoral dissertation).

[16] K. K. Kesari, R. Soni, Q. M. S. Jamal, P. Tripathi, J. A. Lal, N. K. Jha, ... & J. Ruokolainen (2021). Wastewater treatment and reuse: a review of its applications and health implications. Water, Air, & Soil Pollution, 232, 1-28.

[17] L. Metcalf, H. P. Eddy, & G. Tchobanoglous (1991). Wastewater engineering: treatment, disposal, and reuse (Vol. 4). New York: McGraw-Hill.

[18] R. J. Mislevy (1991). [Review of Statistical Analysis with Missing Data, by D. B. Rubin & Roderick J. A. Little]. Journal of Educational Statistics, 16(2), 150–155. https://doi.org/10.2307/1165119

[19] S. Moritz (s. d.). Comparison of different Methods for Univariate Time Series Imputation in R.

[20] D. A. Newman (2014). Missing Data : Five Practical Guidelines. Organizational Research Methods, 17(4), 372-411. https://doi.org/10.1177/1094428114548590

[21] M. Quintero & A. Leboulluec (2018). Missing Data Imputation for Ordinal Data. International Journal of Computer Applications.

[22] L. Rieger, S. Gillot, G. Langergraber, T. Ohtsuki, A. Shaw, I. Takács, S. Winkler, & Guidelines for Using Activated Sludge Models; IWA Publishing; 2012.

[23] D. Wang, M. Ha, & J. Qiao (2021). Data-Driven Iterative Adaptive Critic Control Toward an Urban Wastewater Treatment Plant. IEEE Transactions on Industrial Electronics, 68(8), 7362-7369. https://doi.org/10.1109/TIE.2020.3001840

[24] WEF (Water Environment Federation) & ASCE (American Society of Civil Engineers); Design of Municipal Wastewater Treatment Plants; Fifth Edition; McGraw Hill; 2010; 147-149

[25] Wik & Lindeborg. (1994). Modelling the dynamics of a trickling filter for waste water treatment. Proceedings of IEEE International Conference on Control and Applications CCA-94, 1035-1040 vol.2. https://doi.org/10.1109/CCA.1994.381371

[26] Y. Zhang & P. J. Thorburn (2022). Handling missing data in near real-time environmental monitoring: A system and a review of selected methods. Future Generation Computer Systems, 128, 63-72.