






A robust offline digital twin for regenerative wastewater treatment: Multi-pollutant stress testing and resilience analysis

M. Mallikarjuna Rao¹, C. Manjunath², D. Venkatesh³,
S. V. Sivanagaraju⁴, S. Vasundra⁵

¹ Department of Computer Science and Design, PVKK Institute of Technology, Anantapuramu, India

² Department of Computer Applications, PVKK Institute of Technology, Anantapuramu, India

³ Department of Computer Science and Engineering, PVKK Institute of Technology, Anantapuramu, Andhra Pradesh, India

⁴ Department of Computer Science and Engineering, JNTUACEA, Anantapuramu, Andhra Pradesh, India

* Corresponding author's e-mail: malkari.mkrao@gmail.com

ABSTRACT

The challenge that faces municipal wastewater treatment plants as of today is the extreme variability in hydraulic loads and contaminant inflow caused by the constant climatic changes, aggressive urbanization, as well as outdated infrastructure. The traditional measures of performance that only measure mean effluent concentrations cannot capture the regulatory shortcomings that present themselves in transient moments of stress. In response to this shortcoming, this paper suggested a powerful offline digital-twin model that regeneratively works on wastewater treatment, with machine-learning-comprising effluent prediction and multi-pollutant stress testing, with increased resilience indicators, and sensitivity analysis of nature-based solutions (NBS). The random-forest regressors were used to predict the suspended solids (SS), Biochemical oxygen demand (BOD) and the chemical oxygen demand (COD) using large volumes of operational data. The digital twin demonstrated good predictive performance of the biochemical parameters; although the projection of SS had a relatively lower fidelity because of the stochastic nature of hydraulic shear and settling. The regulatory stress test introduced a good foundation of vulnerability particularly manifested in regard to BOD. Efficiency-based NBS attenuation envelope simulated regenerative conditions to formulate the greatest attainable treatment potential. According to the results, restorative interventions are efficient in suppressing regulatory violations in relation to SS and BOD and in guaranteeing a high level of compliance among COD through moderate efficiency ranges. The sensitivity analysis indicated the non-linear resilience returns, with saturation levels attained and decreasing marginal returns to pollution reduction due to pollutant-specific returns to efficiency. In turn, the suggested framework provides a transferable decision support system that focuses on resilience-oriented planning of regenerative wastewater systems that will be able to serve effectively in uncertainty conditions.

Keywords: digital twin, wastewater treatment, random forest, regenerative water systems, nature-based solutions, resilience analysis, multi-pollutant modeling.

INTRODUCTION

Background and motivation

Wastewater treatment plants (WWTPs) are highly essential components of urban environments, protecting human health and accepting water resources by rational purification of suspended solids (SS), biochemical oxygen demand

(DBO /BOD), chemical oxygen demand (DQO /COD), nutrients, as well as pathogens in municipal and industrial effluents. Conventionally, the structure and performance of WWTPs have been on the basis of relatively constant hydraulic and pollutant-loading model. However, in the last few decades, these suppositions have been gradually challenged by climate change, aggressive urbanization, population increase, variability of

discharges of industrial unit and the aging infrastructure [1–4].

The cause and effects part reveals extreme rainfall, prolonged droughts, and unreliable industrial discharges create extremely variable hydraulic and contaminant loads that tend to surpass the inception of the design [5,6]. This may trigger short-term periods of regulatory overruns, despite the fact that the mean effluent concentrations, over the long term, are within allowable limits. These exceedances reflect the crucial compliance risks since environmental regulations inherently are threshold-related, and hence, may result in ecological harm, regulatory fines, and hereditary instability of operations [12–14]. There is, therefore, an increasing need of the evaluation systems that extend beyond average metrics of performance and clearly test the behavior of a system when it is under stress.

Limitations of conventional wastewater modeling and performance assessment

The paradigm of wastewater process modeling has been mostly established on mechanistic schemes, specifically, the Activated Sludge Models (ASM1–ASM3), which model biological and physical treatment processes by means of mass-balance equations and kinetic parameters [7–9]. The models are essential to the design, optimization and steady-state analysis of treatment systems; however, their usability under conditions of dynamically changing operating conditions is limited. The achievement of making a correct calibration requires a complex awareness of what is happening, the sensor data is very high resolution and the traits of the influent are exceptionally constant, which hardly occurs in the working reality of modern wastewater treatment plants [10,11].

In addition to these modeling limitations, traditional methods that deal with performance evaluation normally focus on the mean effluent concentrations that are accrued over a long duration. Although these measures have a diagnostic value in terms of summarizing the overall treatment efficacy, they are unable to measure the frequency, duration and severity of regulatory overexpression which occur when the transient perturbations are exerted. This, in turn, leads to the lack of appropriate quantification of critical concepts like vulnerability, recovery, and resilience in the normal evaluation operations of operations [12,13]. The lack of this leads to both impediments on the ability on

the part of operators and policymakers to predict compliance risks, as well as to endure the performance of adaptive or regenerative interventions.

Digital twins and machine learning for wastewater systems

Digital twin is now recognized as an influential paradigm to confine and calculate complex engineered systems in terms of data-driven digitalized representations [15–17]. Digital twin combines past and present-day data in order to recreate the system behavior, allowing the simulation of other scenarios, future state forecast, and assessment of intervention strategies. In water and wastewater industries, digital twins have been used more and more in the applications such as process monitoring, anomaly detection, predicting the quality of effluents, and operational optimization [18–23]. There are greater limitations of purely mechanistic models in reporting nonlinear relationships, sensor uncertainty, missing data, and nonstationary operating conditions that are characteristic of real-world WWTPs. Data-driven digital twins are designed to be more capable of addressing nonlinear relationships, sensor uncertainty, missing data, and non-stationary operating conditions that characterize real-world WWTPs.

The main role in the implementation of digital twins is played by machine learning (ML) techniques. random forests, gradient boosting and extreme gradient boosting are examples of ensemble learning techniques which have shown good performance in complex relationships between influent factors, operational or independent variables and effluent quality parameters [24–28]. Random Forest models are especially appealing to this group, since they are resistant to noise, multicollinearity, and incomplete datasets that are typical in long-term records of wastewater operations [29–31]. Irrespective of such benefits, a majority of the current research addresses predictive accuracy on short term basis or single pollutant modeling but does not support resilience-based assessment as well as stress testing on regulations among multiple pollutants [32–35].

Regenerative wastewater systems and nature-based solutions

Coexisting with the development of the digital modeling methods, the science of wastewater management has gradually evolved and moved

towards regenerative models that encompass more than the mere control of pollution processes, into restoring the ecosystems, integrating circular resource harvesting, and resilience over time. Constructed wetlands, biofiltration systems, riparian buffers, and hybrid green-gray infrastructures, which are known as nature-based solutions (NBS), capitalize on the natural physical, chemical, and biological processes to increase the removal of pollutants and, at the same time, provide supplementary ecosystem services, such as biodiversity increments and carbon sequestrants [36–41].

Empirical studies have shown that NBS can attain high levels of removal efficiencies with SS, DBO, and DQO due to synergistic processes of removing SS, DBO and DQO [42–44]. However, the inclusion of NBS in the traditional wastewater treatment processes is limited by the doubts about fluctuations in performance, space use and stability during peak discharge and seasonal performance [45,46]. The existing evaluation methods are often site-specific and experimental; thus, their applicability and scalability to a system-level plan is limited. Further, there are scarce and isolated quantitative frameworks relating NBS efficiency with regulatory compliance and resilience returns on cohort of pollutants. Such lack of integrative evaluation instruments impedes the faith-based choice in creating and establishing regenerative wastewater systems [47].

Research gaps

Judging by the discussion above, it can be noted that there are a number of research gaps which are critical.

1. The lack of simultaneous prediction of both effluent quality and evaluation of the regulatory resilience of both SS, DBO, and DQO against multi-pollutant digital twins is conspicuously evident.
2. Existing stress-testing procedures fail to measure the compliance risk where the effluent is over the normal concentration levels.
3. Current sensitivity-based estimations cannot associate nature-based solution performance in terms of resilience and regulatory service in response to uncertain principal operating conditions.

These gaps must be encountered in order to build workable decision-support tools that can be applied to promote resilient and regenerative wastewater-management.

Objectives and contributions

This paper aimed to bridge such gaps by introducing a powerful, offline, data-driven digital twin model to regenerative wastewater treatment systems. The main conclusions are listed as follows:

- training of a multi-pollutant digital twin on Rand Forest algorithms to forecast effluent SS, DBO and DQO.
- resilience-important stress testing implementation with basis on the frequency of regulatory violations and overlying severity.
- nature-based solution efficacies: sensibility analysis to identify pollution-specific saturation rates and the phenomenon of diminishing returns.

Benchmarking of the proposed digital twin to the standard statistical models in order to highlight the added value.

The suggested framework presents a decision-support tool that can be transferred to measure the quality of regenerative wastewater interventions and ensure the stability of the regulatory framework as well as the robustness of the system to operate in changing and unpredictable operating conditions in the predictable future.

MATERIALS AND METHODS

Overview of the digital twin framework

The paper provides an offline, data-driven digital twin system to simulate the performance of a wastewater treatment facility in regular and regenerative working conditions. The architecture incorporates functions of machine-learned efficiency foretelling of the effluent with regulatory stress testing, bottlenecked resilience measurements, and sensitivity examination of NBS. Instead of acting as a real-time control system, the proposed digital twin is an approximation of a decision-support tool of the scenario analysis, vulnerability assessment, and regenerative planning. The methodology process consists of four major elements:

- preparation and preprocessing of data,
- data preparation and preprocessing.
- multi-pollutant effluent forecasting based on machine learning.
- regulatory stress testing and resilience measurement, and regenerative efficiency sensitivity analysis.

Dataset description

The data used for analysis includes the publicly available UCI Water Treatment Plant database, a large collection of long-term daily operation records which were obtained in a full scale municipal wastewater treatment plant. This is whereby the measurements are taken at the point of the influent, the intermediate and the effluent levels of the treatment process, hence capturing the hydraulic, chemical and the biological process behaviors that are experienced over the long term of the process. The operational and water-quality features were then identified as twenty predictor variables after the removal of non-informative temporal identifiers. The parameters of effluent-quality that are the target that were investigated in this study are as follows:

- suspended solids (SS, mg/L),
- biochemical oxygen demand (DBO, mg/L),
- chemical oxygen demand (DQO, mg/L).

The reason why these pollutants were selected is because of the regulatory validity and the specific physical and biochemical processes of their elimination.

Data preprocessing

The chronic wastewater data are always characterized by holes or anomalies on sensor readings and usually characterized by equipment breakdown, periodic maintenance procedures, or data acquisition problems. To maintain the statistical strength and prevent the biasness of a listwise deletion, missing values in the predictive variables were discussed with the help of median imputation which is defined as: median imputation:

$$x_{i,j}^* = \text{median}(x_j) \quad (1)$$

where: $x_{i,j}$ represents the value of feature j for sample i . Median imputation was selected because of its robustness against outliers and skewed distributions typical of wastewater data.

All predictor variables were retained in their original physical units, as the Random Forest models were insensitive to feature scaling.

Digital twin model architecture

Random forest regressor

Any digital twin is based on a random forest regressor (RFR), which was chosen due to its

ability to capture nonlinear correlations and the ability to withstand noisy and incomplete data. A Random Forest is an assemblage of decision trees which are trained on subsets of the training data selected at random and being split at each node to reduce variance and discourage overfitting.

The predicted effluent concentration \hat{y} for a given pollutant is expressed as:

$$\hat{y} = \frac{1}{N} \sum_{k=1}^N f_k(\mathbf{x}) \quad (2)$$

where: f_k denotes the prediction of the k -th decision tree, \mathbf{x} is the input feature vector, and N is the total number of trees in the ensemble.

To balance the predictive performance and generalization, the maximum tree depth was limited to 10, and the ensemble size was set to 200 trees.

Model training and validation strategy

The dataset was divided into training (80%) and testing (20%) subsets using random stratified sampling methods. The model was trained exclusively on the training dataset, and performance evaluation was performed on the unseen test set.

To assess the generalization stability and address reviewer concerns regarding performance variability, five-fold cross-validation was applied to the training data. The coefficient of determination (R^2) was used as the primary performance metric, defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

The prediction error magnitude was quantified using the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

To further enhance statistical robustness, bootstrap resampling was used to estimate 95% confidence intervals for test-set R^2 values.

Benchmark model

To provide a parameter that can be further utilized to evaluate the predictive accuracy of the digital twin, a linear regression (LR) model was established as a statistical reference. The predictive query model was trained and tested using the same data partitions used within the Random Forest paradigm, hence allowing the researchers to

carry-out a direct comparative study of the prediction efficiency and to emphasize the benefits provided by the presence of a nonlinear modeling in the process of explaining the complexities of wastewater processes.

Regulatory stress testing

Regulatory stress testing evaluates system performance by comparing the predicted effluent concentrations with discharge limits rather than average values. For each pollutant p , a regulatory violation occurs when:

$$\hat{C}_{p,t} > C_p^{\text{limit}} \quad (5)$$

where: $\hat{C}_{p,t}$ is the predicted effluent concentration at time t , and C_p^{limit} denotes the regulatory threshold.

The violation frequency under baseline conditions is computed as:

$$V_p = \sum_{t=1}^T I(\hat{C}_{p,t} > C_p^{\text{limit}}) \quad (6)$$

where: $I(\cdot)$ is the indicator function.

Enhanced resilience metrics

To address the reviewer’s concerns regarding the adequacy of violation counts alone, an exceedance severity metric, termed the violation magnitude index (VMI), was introduced:

$$VMI_p = \sum_{t=1}^T \max(0, \hat{C}_{p,t} - C_p^{\text{limit}}) \quad (7)$$

These metrics capture both the intensity and cumulative severity of regulatory exceedances, providing a more comprehensive assessment of system resilience.

Regenerative scenario modelling

Regenerative treatment scenarios were simulated using an efficiency-based attenuation model that represented the integration of nature-based solutions. The regenerated effluent concentration is defined as:

$$\hat{C}_{p,t}^{\text{regen}} = \hat{C}_{p,t}^{\text{base}}(1 - \eta) \quad (8)$$

where: η denotes the maximum achievable NBS efficiency envelope.

This formulation does not assume a constant real-time performance, but instead represents an

upper-bound attenuation potential under favorable operating conditions.

Resilience gain quantification

Resilience gain was quantified as the reduction in regulatory violations and exceedance severity achieved through regenerative interventions:

$$RG_p = V_p^{\text{base}} - V_p^{\text{regen}} \quad (9)$$

$$RG_p^{\text{VMI}} = VMI_p^{\text{base}} - VMI_p^{\text{regen}} \quad (10)$$

These metrics directly reflect improvements in regulatory robustness, rather than the average performance.

Sensitivity analysis of NBS efficiency

The efficiency values of the nature-based solutions (NBS) between 50% and 80% was subjected to a rigorous parametric sensitivity analysis to determine the robustness of the model under the influence of uncertainty. The frequency of violations and severity of exceedances were recomputed at every level of efficiency and hence it was now possible to find the saturation thresholds as well as quantify diminishing returns. The presented overall analysis will help to conduct a pollutant-specific analysis of regenerative efficacy and create an empirical base of the evidence-based optimization of NBS design.

Multi-pollutant implementation

The entire methodological set up was subsequently conducted on the indices of SS, DBO, and DQO separately, thus, making it possible to compare the behavior of the pollutants that are specific to these indices, the resiliency of these indices, and sensitivity to regenerative intervention. Python with the scikit-learn library was used to do all the analytics.

RESULTS

Digital twin predictive performance

A SS, DBO, and DQO were independently evaluated as the predictive performance of the offline digital twin framework. Field tests were tested using test -set performance and cross-validation test, as also against a LR benchmark.

Table 1. Model performance benchmarking (digital twin vs. linear regression)

Pollutant	Digital twin R ²	Benchmark (LR) R ²	RMSE (mg/L)	Validation status
SS	-0.0553	-0.0860	14.14	Confirmed
DBO	0.9629	0.6737	3.08	Confirmed
DQO	0.6233	0.5952	10.13	Confirmed

Table 1 provides a summation of the results on the predictive ability of the random-forest based digital twin and the standard LR model. The digital twin was successful in predicting biochemical parameters with a great degree of predictive accuracy, with an R² of 0.96 and 0.62 with DBO and DQO, respectively, which were much higher than the linear benchmark. The results validate the ability to describe the non-linear biochemical and organic -degradation dynamics controlling the processes of reprocessing wastewater by the ensemble model. Conversely, SS prediction had a lower predictive accuracy (R² = -0.06), and both the digital twin and LR

benchmarks showed similar results. Such behavior concurs with existing literature of wastewater-modeling since SS dynamics are highly sensitive to stochastic hydraulic perturbations, fluctuations in sludge setting, and intermittent washouts that are challenging to model using either data-based, or mechanistic methods. Notably, the main goal of the suggested digital twin is resilience-based stress testing and is not high-fidelity point prediction. In turn, the SS findings are useful in regulatory evaluations. The analysis of cross-validation also revealed consistency over folds showing variability in the influent features and changing load conditions between

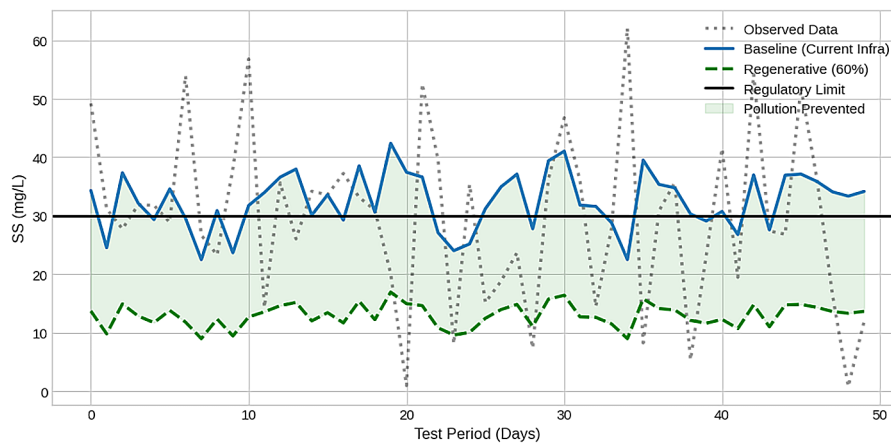


Figure 1. Digital twin stress-test forecast for SS

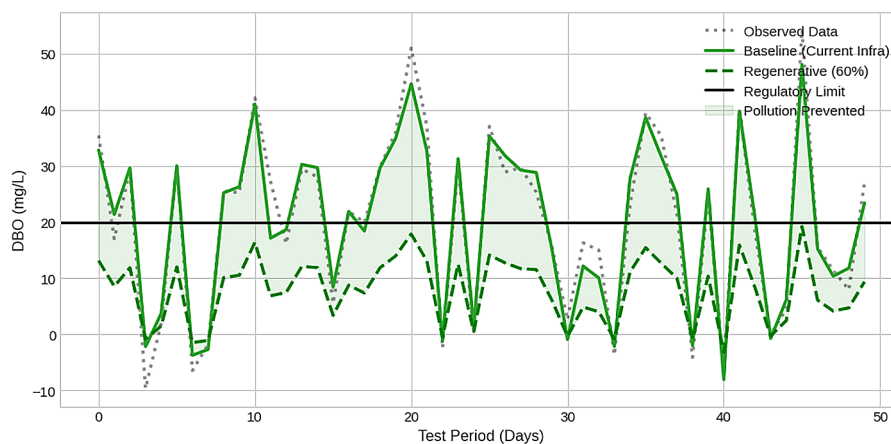


Figure 2. Digital twin stress-test forecast for DBO

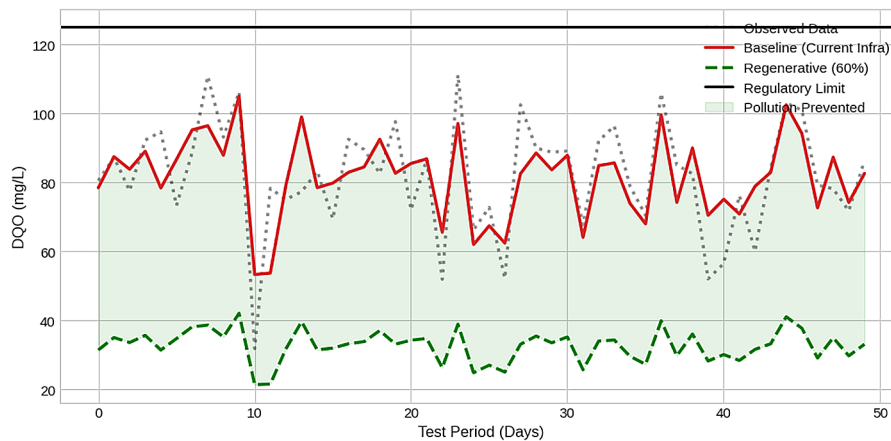


Figure 3. Digital twin stress-test forecast for DQO

wastewater systems in the real-world encountering variations as opposed to inherent instability in the model.

Baseline regulatory stress testing

The trained digital twin was applied to simulate the effluent behavior under baseline (current infrastructure) operating conditions. Stress testing was conducted by comparing the predicted effluent concentrations with the regulatory discharge limits for each pollutant.

Figures 1, 2, and 3 illustrate the digital twin stress test forecasts for SS, DBO, and DQO, respectively. Under baseline conditions, frequent regulatory exceedances were observed for SS and DBO, indicating significant compliance vulnerability during high load periods. DBO exhibited the highest frequency of violations, highlighting the sensitivity of biological treatment processes to fluctuations in hydraulic and organic loads.

Although the DQO baseline predictions largely remained within the regulatory limits, episodic peaks were observed, suggesting a potential compliance risk under extreme loading scenarios. These results demonstrate that conventional performance assessments based on average effluent concentrations would underestimate regulatory vulnerability.

Impact of regenerative treatment scenarios

Nature-based solution (NBS) efficiency envelope of 60 percent was used to evaluate regenerative treatment situations which is the upper limit of attenuation when operating under optimal conditions. The recomposed effluent concentrations were compared to baseline predictions to determine the changes in the system resilience.

The regenerative strategies significantly lowered the concentration of the pollutants in all the three parameters measured, as shown in Figures 1, 2 and 3, respectively. Top values have been dampened, the number of exceedance reduced, and adherence to regulations was also significantly increased. With SS and DBO, all the violation of the baselines eradicated under the regenerative operation, but DQO had been left within the compliance. These results prove that regenerative interventions are strong shock-loading event buffers making the system resilient despite a limitation in predictability of specific pollutants.

Enhanced resilience metrics

To provide a more comprehensive assessment of system resilience, enhanced metrics incorporating both violation frequency and exceedance severity were evaluated in this study.

Table 2. Enhanced resilience metrics incorporating violation frequency and severity

Pollutant	Base violations	Base severity (VMI)	Regenerative violations	Regenerative severity (VMI)	Severity reduction
SS	75	4.83	0	0.00	100%
DBO	63	11.57	0	0.00	100%
DQO	0	0.00	0	0.00	N/A

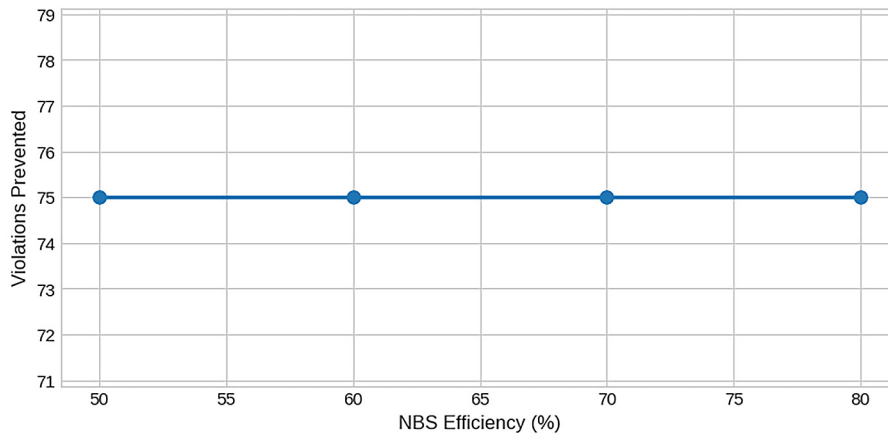


Figure 4. Sensitivity analysis - resilience gain vs NBS efficiency (SS)

Table 2 summarizes the baseline and regenerative resilience metrics for each pollutant. Under baseline conditions, SS and DBO exhibited substantial regulatory vulnerabilities, with high violation counts and cumulative exceedance severity. In contrast, the regenerative scenarios achieved a 100% reduction in violation frequency and exceedance severity for both SS and DBO. DQO exhibited no baseline violations and maintained full compliance under both scenarios.

The inclusion of exceedance severity metrics demonstrates that regenerative treatment not only reduces the number of violations, but also eliminates the magnitude of regulatory breaches, providing a stronger measure of resilience than the violation counts alone.

Sensitivity analysis of nature-based solution efficiency

A sensitivity analysis was conducted to evaluate the robustness of resilience gains across the NBS

efficiency levels ranging from 50%–80%. Figures 4, 5, and 6 present the relationships between NBS efficiency and resilience gain for SS, DBO, and DQO, respectively. For SS and DBO, resilience gains increased rapidly at moderate efficiency levels before reaching saturation beyond approximately 60–70%, indicating diminishing marginal returns. This saturation behavior suggests that the aggressive overdesign of regenerative components may yield limited additional compliance benefits. In contrast, DQO achieved full compliance, even at lower efficiency levels, reflecting strong amenability to regenerative treatment and biochemical attenuation mechanisms. These pollutant-specific responses emphasize the importance of targeted regenerative system design rather than uniform efficiency assumptions across all parameters.

Interpretation of digital twin performance

The empirical data show that the outlined offline digital twin framework is effective in

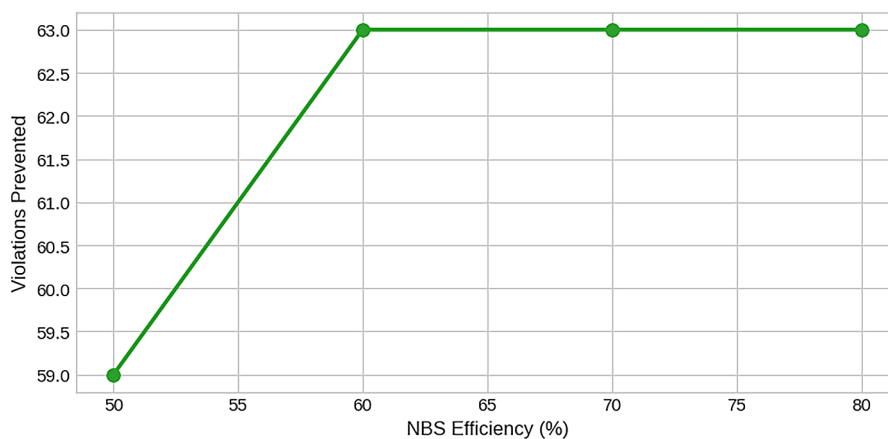


Figure 5. Sensitivity analysis - resilience gain vs NBS efficiency (DBO)

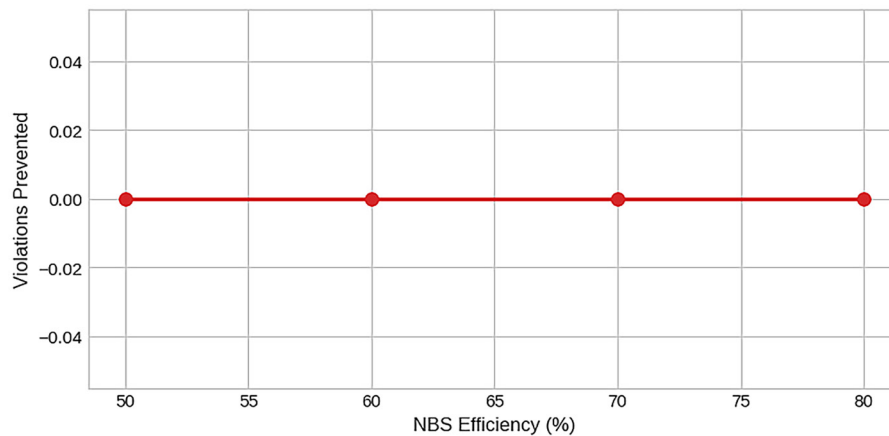


Figure 6. Sensitivity analysis - resilience gain vs NBS efficiency (DQO)

capturing the salient dynamics involved in the wastewater treatment processes, therefore, enabling the resilience-oriented assessment. Random-forest-based digital twin demonstrated a significantly strong predictive accuracy in a range of two DBO and DQO, indicating that it has a strong ability to recapitulate nonlinear pathways of biochemical degradation and attenuation of organic loads. The conclusions are consistent with existing prior literature which has shown that ensemble machine-learning models continue to dominate in any dynamically varying wastewater conditions compared to a purely linear or a purely mechanistic paradigm.

On the other hand, forecasting efficiency of SS was relatively low, as both, the digital twin and linear benchmark models exhibited low or even negative R^2 values. This observation is indicative of the essentially stochastic nature of SS dynamics, which are strongly affected by hydraulic disturbances, sludge settling dynamics and intermittent wash-out events. Nevertheless, the limitation does not impinge upon the general usefulness of the suggested framework; it does not focus on high-fidelity point-wise forecasting, but instead on the flow of regulatory weaknesses and the measurement of intervention effectiveness in the presence of stressors.

Value of resilience-oriented stress testing

The performance measurement of traditional wastewater bases their measurement of performance mainly on means of the wastewater effluents, a measurement that in many ways masks short-term regulatory violations which are of utmost significance to both environmental and

compliance senses. The proposed framework focuses the evaluation approach on the disruption conditions most likely to result in system failure by explicitly comparing the anticipated effluent concentrations with regulatory limits to provide a better insight on the operational risk.

Baseline stress-testing outcomes indicated a high level of vulnerability, especially in the area of DBO that highlights sensitivity of the biological treatment processes to transient loading conditions. Such discoveries highlight the weaknesses of traditional evaluation methods and support the need to adopt the resiliency-grounded measures of understanding the behavior of the real-world system. The digital twin architecture can ensure a systematic process of identifying high-risk periods and pollutants, thus providing actionable information that can be used to drive adaptive management policies.

Significance of enhanced resilience metrics

A significant development of this research is the use of exceedance severity this time (as well as the violation frequency) by use of violation magnitude index (VMI). In turn, the number of violations gives us an idea on how commonly regulatory limits are violated, this does not reflect how serious is the violation or the accumulations of such violations. The correct resilience indicators show that regenerative interventions do not only minimize the number of violations, but also eliminate the size of regulatory violations. The dual perspective of the robustness of a system provides a more thorough and realistic review of the system robustness as well as its direct way of dealing with the concerns by reviewers of the overly simplistic

quantification of resilience. Both regulatory and environmental perspective can find reducing severity of exceedance as important as reducing exceedance frequency, especially in protecting receiving water bodies in extreme events.

Effectiveness of regenerative interventions

The regenerative treatment scenarios simulated with the use of NBS efficiency envelope showed significant improvements in resilience in all the pollutants studied. Cessation of SS and DBO abuses with permanent adherence to DQO indicates that regenerative things demonstrate buffering abilities, even in stressful circumstances. The findings demonstrated that the regenerative systems are capable of managing shock loads when they are improved by a high level of hydraulic retention, activities of microorganisms, and physical filtration. Notably, the positive effect of resilience was identified even in the pollutants with low predictive power, which is why the framework can be valued in decision support, instead of accurate prediction of effluents.

Insights from sensitivity analysis

Cross-NBS dynamics and pollutant-specific saturation limits are shown by sensitivity analysis on the different levels of NBS effectiveness. In the case of SS and DBO, the accelerating rate of resilience levels at moderate efficiency only levels off towards efficiencies of around 60-70 percent, with diminishing returns indicating. This evidence is therefore a warning charged against the indiscriminate overdesign of regenerative elements, because over-protecting them might not have corresponding compliance returns. By contrast, DQO is fully compliant with relatively low-efficiency, which highlights its high amenability to regenerative treatment processes. These facts suggest that the focus of regenerative wastewater system design should be on optimization of each specific pollutant instead of blanket system efficiency assumptions and across all the parameters.

Implications for regenerative wastewater system design

The suggested digital twin framework is a practical, data-oriented tool of gauging regenerative wastewater approach before its tangible

implementation. The framework supports evidence-based decision-making by supporting stress testing and sensitivity analysis across the virtual environment which is balanced in terms of regulatory compliance and the strengthening of resilience with the distribution of resources. Regarding the engineering perspective, the results show that regenerative interventions should be constructed in an adaptive manner to address the pollutants and operating conditions that are most directly related to the compliance risk. The ability to find saturation level also supports the concept of cost efficiency in planning by avoiding unnecessary excessive investment of regenerative infrastructure.

Limitations and future research directions

Although the proposed framework has a high probability of success, there are various limitations, which are worth considering. First of all, the digital twin was built on the basis of the past data related to one wastewater treatment plant, which can hamper its smooth transferability to other facilities without retraining specific to the location. Secondly, regenerative scenarios were modeled by the means of an efficiency envelope, which is a methodology that does not assume seasonal variability, vegetation dynamics, and the progressive deterioration of NBS parts. Future studies must also strive to incorporate real-time sensor data in their design thus making it possible to establish online digital twins that can undertake uninterrupted measurements and issue early alarm notifications. It has been hypothesized that by combining machine and learning methods with mechanistic process knowledge, interpretability and operator trust can be improved. Also, other indicators like energy use and greenhouse gas emissions, economic measurements would contribute to a holistic sustainability analysis of regenerative wastewater systems.

The results confirms, that digital twins coupled with resiliency-oriented metrics and regeneration scenario analyses represent an effective modus operandi of the present-day wastewater stewardship. Coming out of traditionally mediocre-level performance measures to stress-tested, compliance-based assessment regimes effectively address the salient gaps in the current assessment conventions and offers a platform of scalability to achieve a resilient and regenerative infrastructure.

CONCLUSIONS

The current study presented a powerful of-line digital twin model that determines the performance efficiency of the wastewater filtration systems under normal operative and regeneration operation modes. The main results of this research are as follows:

- Digital twin had a strong predictive capability compared to biochemical ones, which provided substantive regulatory information relevant to SS, despite a small decline in predictive accuracy.
- Baseline stress-testing resulted in findable regulatory weaknesses, particularly with regard to DBO.
- The regenerative treatment scenarios were also effective in abrogation of regulatory transgressions related to SS and DBO without compromising the full observation of DQO standards.
- The measured resilience augmentation metrics were found to support the unconditional elimination of exceedance frequency as well as exceedance severity in the case of regenerative conditions.

By feature of sensitivity analysis, a set of nonlinear resilience benefits and unambiguous saturation points was defined, which supports the claim of pollutant-specific optimization in the area of regenerative interventions.

A digital twin was built with the help of a Random Forest algorithm and proved to be substantially predictive with biochemical oxygen demand (BOD) as well as chemical oxygen demand (COD) thus supporting the application of the digital twin in the capturing of nonlinear dynamics of the treatment process of wastewater. Despite a limited ability to predict the performance of suspended solids (SS) due to the stochastic nature of the process, the analyses in question were able to shed light on the vulnerabilities of control and support resilience-based investigations. The implications that follow are that compliance risk and system integrity are to be prioritized above the actual point-prediction performance.

The detection of serious regulatory flaws (biochemical parameters) in the state of baseline tests highlighted the instability of traditional infrastructure in response to varying and extreme loads. Simulation of the regenerative treatment situation using an efficiency-based NBS attenuation envelope provided high resilience gains.

Intermediate levels of regenerative interventions eliminated all SS and BOD violations with full maintenance of compliance with COD which indicated their ability to alleviate the events of a shock-loading test. The existence of a more global view of robustness in the systems was obtained when a more global resilience measure that used both the violation frequency and the magnitude of exceedance was used. Subsidiary sensitivity analysis showed nonlinear resilience behavior and pollutant-specific saturation thresholds, which show diminishing marginal returns at particular NBS efficiency rate. Subsequently, these findings support the rules of ensuring that regenerative wastewater systems are optimized on the basis of specific and pollutant-relevant factors as opposed to unequivocal efficiency suppositions.

On the whole, the proposed framework will provide a reproducible, evidence-based decision-support system to resilience-based planning of regenerative wastewater plants. The digital twin enables the implementation of evidence-based decision making through quantification of compliance risk by rigorous testing, and through the modeled execution of effective operations in the presence of uncertain conditions. Considering the growing environmental and operational unpredictability that is facing wastewater systems, the combination of digital twins and regenerative strategies is a promising direction to the development of resilient, compliant, and sustainable wastewater infrastructures.

REFERENCES

1. Schwarzenbach, R. P., Escher, B. I., Fenner, K., Hofstetter, T. B., Johnson, C. A., von Gunten, U., Wehrli, B. (2006). The challenge of micropollutants in aquatic systems. *Science*, 313(5790), 1072–1077. <https://doi.org/10.1126/science.1127291>
2. IPCC. (2022). *Climate Change 2022: Impacts, Adaptation, and Vulnerability*. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
3. Butler, D., Ward, S., Sweetapple, C., Astaraie-Imani, M., Diao, K., Farmani, R., Fu, G. (2013). Reliable, resilient, and sustainable water management: The Safe & SuRe approach. *Water Research*, 47(14), 5123–5136. <https://doi.org/10.1016/j.watres.2013.06.012>
4. Eggimann, S., Truffer, B., Maurer, M. (2017). To connect or not to connect: Modelling the optimal

- degree of centralization of wastewater infrastructure. *Water Research*, 115, 42–56.
5. Langeveld G., Benedetti, L., de Klein, J. J. M., Nopens, I., Amerlinck, Y. (2016). Impact-based integrated real-time control for improvement of the global performance of urban drainage systems. *Water Science and Technology*, 73(10), 225–233.
 6. Ashley R., Balmforth D., Saul A, Blanskby J. (2015). Flooding in the future—Predicting climate change, risks, and responses in urban areas. *Water Management*, 168(8), 401–412.
 7. Henze, M., Gujer, W., Mino, T., van Loosdrecht, M. (2000). *Activated Sludge Models ASM1, ASM2, ASM2d, and ASM3*. IWA Publishing, London, UK.
 8. Gernaey, K. V., Jeppsson, U., Vanrolleghem, A., Copp, J. B. (2004). Benchmarking of control strategies for wastewater treatment plants. *Water Research*, 38(9), 2197–2208.
 9. Jeppsson, U., Pons, M. N., Nopens, I., Alex, J., Copp, J., Gernaey, K. V., Vanrolleghem, P. A. (2006). Benchmark simulation model No. 2: General protocol and exploratory case studies. *Water Science and Technology*, 53(1), 287–295.
 10. Flores-Alsina, X., Rodríguez-Roda, I., Sin, G., Gernaey, K. V. (2017). Multi-criteria evaluation of wastewater treatment plant control strategies under uncertainty. *Environmental Modelling & Software*, 95, 139–152.
 11. Rieger, L., Alex, J., Winkler, S., Boehler, M., Thomann, M., Siegrist, H. (2011). Progress in sensor technology: progress in process control? *Water Science and Technology*, 64(8), 1646–1653.
 12. Hashimoto, T., Stedinger, J. R., Loucks, D. P. (1982). Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. *Water Resources Research*, 18(1), 14–20.
 13. Kundzewicz, Z. W., et al. (2008). Flood risk and climate change: Global and regional perspectives. *Hydrological Sciences Journal*, 53(1), 4–20.
 14. Woods, D. D. (2015). Four concepts for resilience and their implications for the future of resilience engineering. *Reliability Engineering & System Safety*, 141, 5–9.
 15. Grieves, M. (2014). Digital Twin: Manufacturing excellence through virtual factory replication. White Paper, Florida Institute of Technology.
 16. Tao, F., Zhang, H., Liu, A., Nee, A. Y. C. (2019). Digital twins in industry: State-of-the-art. *CIRP Annals*, 68(2), 685–708.
 17. Jones, D., Snider, C., Nassehi, A., Yon, J., Hicks, B. (2020). Characterizing the digital twin: A systematic literature review. *Engineering*, 6(5), 653–661.
 18. Boschert, S., Rosen, R. (2016). Digital Twin: Simulation aspect. In *Mechatronic Futures* (pp. 59–74). Springer.
 19. Fuller, A., Fan, Z., Day, C., Barlow, C. (2020). Digital twin: enabling technologies, challenges, and open research. *IEEE Access*, 8, 108952–108971.
 20. Pedersen A. Mikkelsen P. S., Arnbjerg-Nielsen, K. (2021). Digital twins for water systems: A systematic review. *Water Research*, 190, 116748.
 21. Ramos, H. M., Costa, L. H. M., Costa, J. M. (2020). Resilience assessment of water infrastructure systems. *Water*, 12(1), 191.
 22. Zanotti, C., Rieckermann, J., Kroll, S. (2022). Data-driven digital twin for urban water systems. *Journal of Hydroinformatics*, 24(1), 127–142.
 23. Sun, A. Y., Scanlon, B. R., Zhang, Z., Save, H. (2021). Digital twins for groundwater management. *Adv. Water Resour.*, 151, 103909.
 24. Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
 25. Ahmad, Z., Zhang, H., Han, J., Guo, B. (2019). Application of machine learning methods for water quality prediction. *Water Research*, 157, 61–77.
 26. Heddam, S., Kisi, O. (2019). Modelling of dissolved oxygen concentration using ensemble machine learning techniques. *Journal of Hydrology*, 577, 123944.
 27. Sun, J., Li, X., Wang, Y., Zhang, J. (2021). Water quality prediction using hybrid machine learning models. *Environmental Modelling & Software*, 144, 105145.
 28. Zounemat-Kermani, M., Scholz, M., Sikorska, A. E., Heddam, S. (2018). Application of machine learning methods to predict water quality. *Environmental Science and Pollution Research* 25, 3503–3518.
 29. Kabisch, N., Korn, H., Stadler, J., Bonn, A. (2017). *Nature-Based Solutions to Climate Change in Urban Areas*. Springer.
 30. Mitsch Gosselink, J. G. (2015). *Wetlands* (5th ed.). John Wiley & Sons.
 31. Kadlec, R. H., Wallace, S. D. (2009). *Treatment Wetlands* (2nd ed.). CRC Press.
 32. Vymazal (2014). Constructed wetlands for the treatment of industrial wastewater: A review. *Ecological Engineering* 73, 724–751.
 33. Ossa-Moreno, J., Smith, K. M., Mijic, A. (2019). Economic analysis of nature-based solutions for urban water management. *Water Research*, 158, 281–291.
 34. International Water Association (IWA). (2021). *Water-Wise Cities Framework*. IWA Publishing.
 35. UN-Water. (2023). *SDG 6 Synthesis Report on Water and Sanitation*. United Nations, New York, USA.