Machine Learning for Predicting SuDS Performance under Climate Extremes

Apprentissage automatique pour prédire la performance des techniques alternatives face aux extrêmes climatiques

Dónoban-S. Rojas-Céspedes[1], Sandra Galarza-Molina[2], Andrés Torres[3]

[1] Maestría en Ciencia de Datos, Pontificia Universidad Javeriana
[2] Departamento Ingeniería Civil e Industrial, Facultad de Ingeniería y Ciencias, Pontificia Universidad Javeriana
[3] Instituto Javeriano del Agua, Pontificia Universidad Javeriana

RÉSUMÉ

Cette étude se concentre sur la modélisation et l'optimisation du système de Artificial Wetland/Reservoir Tank System (SHATR) de l'Université Javeriana en utilisant l'apprentissage automatique afin de renforcer la résilience face aux extrêmes climatiques. Les données historiques de précipitations et de ruissellement (2014–2016) ont été traitées en 98 événements pluvieux pour la modélisation prédictive. Parmi les algorithmes testés, la régression à vecteurs de support (SVR), optimisée par recherche bayésienne, a obtenu la meilleure performance (R^2 = 0,9024 ; RMSE = 1,59 cm). Les principaux prédicteurs incluaient la pluie accumulée, l'intensité maximale et le niveau initial du réservoir. Les scénarios de changement climatique (2021–2100) ont projeté une hausse de 6,1 % des événements de débordement pour 2021–2040 contre 3,1 % historiquement. Bien que la modélisation des sorties ait montré des limites, la prédiction des entrées s'est révélée viable pour l'alerte précoce et la gestion adaptative. L'apprentissage automatique (AA) a démontré son efficacité et sa robustesse pour les relations hydrologiques non linéaires, surpassant les modèles hydrodynamiques traditionnels. Les résultats soulignent la nécessité de stratégies proactives : augmentation de la capacité de stockage, optimisation du déversoir et intégration d'alertes en temps réel. Les travaux futurs devraient inclure un suivi prolongé, des variables supplémentaires et des modèles hybrides. Cette recherche confirme l'AA comme un outil puissant pour la gestion résiliente de l'eau urbaine face à l'incertitude climatique.

ABSTRACT

This study focuses on modeling and optimizing the Artificial Wetland/Reservoir Tank System (SHATR) at Javeriana University using machine learning to enhance resilience against climate extremes. Historical precipitation and runoff data (2014–2016) were processed into 98 rainfall events for predictive modeling. Among tested algorithms, Support Vector Regression (SVR) optimized via Bayesian search achieved the best performance (R^2 = 0.9024, RMSE = 1.59 cm). Key predictors included accumulated rainfall, peak intensity, and initial tank level. Climate change scenarios (2021–2100) projected a 6.1% rise in overflow events for 2021–2040 versus 3.1% historically. While outflow modeling showed limitations, inflow prediction proved feasible for early warning and adaptive management. Machine learning (ML) demonstrated efficiency and robustness for non-linear hydrological relationships, outperforming traditional hydrodynamic models. Findings highlight the need for proactive strategies: increasing storage capacity, optimizing spillway design, and integrating real-time alerts. Future work should incorporate longer monitoring, additional variables, and hybrid models. This research validates ML as a powerful tool for resilient urban water management under climate uncertainty.

KEYWORDS

Sustainable Drainage Systems (SuDS), Machine Learning, Climate Change Adaptation, Hydrological Prediction, Urban Water Resilience

1 INTRODUCTION

Urban water management faces critical challenges due to intensified extreme weather events and rapid urbanization, which increase runoff and overload conventional drainage systems, leading to recurrent flooding with severe socio-economic impacts (Bilgiç & Baba, 2023; Farias & Mendonça, 2022). Sustainable Drainage Systems (SuDS) have emerged as resilient alternatives, mimicking natural processes of infiltration and retention to improve water quality and biodiversity (Fletcher et al., 2015; Nazarpour et al., 2023). However, their performance under climate change scenarios remains uncertain (Allan et al., 2020). Traditional hydrodynamic models, while accurate, are computationally demanding and slow, limiting their use for real-time adaptation (Jin et al., 2022). Machine Learning (ML) offers a promising solution, enabling efficient modeling of complex non-linear hydrological relationships without explicit physical equations (Xu & Liang, 2021). This study applies ML techniques to predict SuDS behavior under projected climate extremes, aiming to strengthen urban resilience and inform adaptive strategies.

2 MATERIALS AND METHODS

The SHATR system at Javeriana University integrates rainwater capture, sedimentation tanks, and a subsurface-flow constructed wetland planted with *Cyperus papyrus* for water quality improvement (Galarza-Molina et al., 2015; Ardila-Quintero et al., 2016). Between 2014–2016, a monitoring network recorded precipitation, water levels, and quality at one-minute intervals, providing a robust dataset for predictive modeling. Historical records were consolidated into a unified DataFrame (1.16 million rows, 12 variables) using Python scripts and manual checks. Quality control excluded incomplete or anomalous events, resulting in 98 valid rainfall events. Key predictors included accumulated rainfall, peak intensity, initial tank level, and event duration. Variables were standardized (e.g., precipitation converted to mm) and selected via Spearman correlation and PCA.

Five machine learning algorithms were tested: Linear Regression, ElasticNet, Random Forest, XGBoost, and Support Vector Regression (SVR). Hyperparameters were optimized using Bayesian search (BayesSearchCV), evaluated with R^2, RMSE, and MAE metrics (Kwon & Kim, 2021; Mosavi et al., 2018).

Future projections (IDEAM, 2021–2100) applied the delta-change method to historical data, adjusting rainfall intensity and duration per Clausius-Clapeyron scaling (Nayak et al., 2020). Four time horizons (2021–2040, 2041–2060, 2061–2080, 2081–2100) were modeled to assess system resilience under extreme conditions. Implementation used Python (pandas, scikit-learn) in Google Colab, ensuring reproducibility and computational efficiency (Xu & Liang, 2021).

3 RESULTS AND DISCUSSION

The study evaluated five machine learning models for predicting inflow levels in the SHATR system. After Bayesian optimization, Support Vector Regression (SVR) achieved the highest accuracy ($R^2 = 0.9024$, RMSE = 1.59 cm), improving by 22% over its baseline ($R^2 = 0.6819$). Tree-based models (Random Forest, XGBoost) also performed well ($R^2 \approx 0.88$), while linear models (ElasticNet, Linear Regression) showed moderate accuracy ($R^2 \approx 0.72$), confirming their limitations for non-linear hydrological relationships (Kwon et al., 2021). Residual analysis validated SVR's robustness and generalization, with errors mostly within ±2 cm and a normal distribution (Shapiro-Wilk $p = 0.77$).

Feature importance analysis revealed accumulated rainfall, peak intensity, and initial tank level as the most influential variables, while humidity and dew point had negligible impact, consistent with Yang & Chui (2021). Applying SVR to IDEAM climate scenarios (2021–2100) projected a 6.1% increase in overflow events for 2021–2040 compared to 3.1% historically, underscoring the need for adaptive measures. Later scenarios (2061–2100) showed reduced overflow frequency but persistent hydraulic stress, aligning with Willems et al. (2012), who predicted intensified short-duration rainfall under climate change.

Outflow modeling was less successful (best $R^2 = 0.25$), due to short monitoring periods and missing variables such as evapotranspiration and substrate retention (Beven, 2009). These limitations highlight the complexity of hydraulic processes and the need for extended monitoring and additional parameters.

4 CONCLUSIONS

This study successfully implemented an optimized Support Vector Regression (SVR) model, achieving high predictive accuracy for inflow levels in the SHATR system ($R^2 = 0.9024$, RMSE = 1.59 cm), validating the potential of machine learning (ML) for modeling non-linear hydrological dynamics. Feature importance analysis highlighted accumulated rainfall, peak intensity, and initial tank level as critical variables, guiding future monitoring strategies (Yang & Chui, 2021).

Climate change scenarios (IDEAM 2021–2100) projected a 6.1% increase in overflow events for 2021–2040 compared to 3.1% historically, underscoring the need for adaptive measures, consistent with Willems et al. (2012). Despite limitations in outflow modeling (best $R^2 = 0.25$), due to short monitoring periods and missing variables such as evapotranspiration (Beven, 2009), the approach demonstrates ML's strategic role in resilient urban water management.

Overall, findings confirm ML's efficiency for modeling complex hydrological dynamics and its potential as an early warning tool for resilient urban water management under climate uncertainty. Recommendations include expanding storage capacity, redesigning spillways, and implementing real-time alerts based on ML predictions. Future work should integrate longer datasets, hybrid models for interpretability, and predictive dashboards to support decision-making. This work reaffirms ML as a key tool for climate adaptation in sustainable urban drainage systems.

5 ACKNOWLEDGMENTS

This work forms part of the initiative *"Adaptation of Sustainable Urban Drainage Systems (SuDS) at PUJ Bogotá/Cali campuses to climate change impacts"* (Project ID 20807), funded under the Multicampus Research Support Program of Pontificia Universidad Javeriana (VRI13-2023).

LIST OF REFERENCES

Allan, R. P., Barlow, M., Byrne, M. P., Cherchi, A., Douville, H., Fowler, H. J., Gan, T. Y., Pendergrass, A. G., Rosenfeld, D., Swann, A. L. S., Wilcox, L. J., & Zolina, O. (2020). Advances in understanding large-scale responses of the water cycle to climate change. *Annals of the New York Academy of Sciences*, *1472*(1), 49–75. https://doi.org/10.1111/nyas.14337

Ardila-Quintero, C., León-Ramírez, R., Galarza-Molina, S., & Torres, A. (2016). Usos potenciales del efluente del humedal-construido en la Pontificia Universidad Javeriana Bogotá. *Revista UDCA Actualidad & Divulgación Científica*, *19*(1), 237–242.

Beven, K. (2009). *Environmental modelling: An uncertain future? An introduction to techniques for uncertainty estimation in environmental prediction.* Routledge.

Bilgiç, E., & Baba, A. (2023). Effect of urbanization on water resources: Challenges and prospects. In *Groundwater in Arid and Semi-Arid Areas: Monitoring, Assessment, Modelling, and Management* (pp. 81–108). Springer.

Farias, A., & Mendonça, F. (2022). The Urban Environmental System perspective on socio-environmental risks of urban flooding. *Sociedade & Natureza*, *34*, e63717. https://doi.org/10.14393/SN-v34-2022-63717

Ferrans, P., Torres, M. N., Temprano, J., & Rodríguez Sánchez, J. P. (2022). Sustainable urban drainage system (SuDS) modeling supporting decision-making: A systematic quantitative review. *Science of The Total Environment*, *806*, 150447. https://doi.org/10.1016/j.scitotenv.2021.150447

Fletcher, T. D., Andrieu, H., & Hamel, P. (2015). Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art. *Advances in Water Resources*, *51*, 261–279. https://doi.org/10.1016/j.advwatres.2012.09.001

Galarza-Molina, S., Torres, A., Lara-Borrero, J., Méndez-Fajardo, S., Solarte, L., & González, L. (2015). Towards a constructed wetland/reservoir-tank system for rainwater harvesting in an experimental catchment in Colombia. *Ingeniería y Universidad*, *19*(2), 169–185.

Jin, L., Kim, M., & Chon, J. (2022). Modeling the resilient supply of ecosystem function for climate change adaptive management in Wetland City. *Journal of Environmental Management*, *322*, 115788. https://doi.org/10.1016/j.jenvman.2022.115788

Kwon, S. H., & Kim, J. H. (2021). Machine learning and urban drainage systems: State-of-the-art review. *Water*, *13*(24), 3545. https://doi.org/10.3390/w13243545

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536. https://doi.org/10.3390/w10111536

Nazarpour, S., Gnecco, I., & Palla, A. (2023). Evaluating the effectiveness of bioretention cells for urban stormwater management: A systematic review. *Water*, *15*(5), 913. https://doi.org/10.3390/w15050913

Nayak, S., & Takemi, T. (2020). Clausius-Clapeyron scaling of extremely heavy precipitations: Case studies of the July 2017 and July 2018 heavy rainfall events over Japan. *Journal of the Meteorological Society of Japan*, *98*(6), 1147–1162.

Willems, P., Arnbjerg-Nielsen, K., Olsson, J., & Nguyen, V. (2012). Climate change impact assessment on urban rainfall extremes and urban drainage: Methods and shortcomings. *Atmospheric Research*, *103*, 106–118. https://doi.org/10.1016/j.atmosres.2011.04.003

Xu, T., & Liang, F. (2021). Machine learning for hydrologic sciences: An introductory overview. *Wiley Interdisciplinary Reviews: Water*, *8*(5), e1533. https://doi.org/10.1002/wat2.1533

Yang, Y., & Chui, T. F. M. (2021). Modeling and interpreting hydrological responses of sustainable urban drainage systems with explainable machine learning methods. *Hydrology and Earth System Sciences*, *25*(11), 5839–5858. https://doi.org/10.5194/hess-25-5839-2021.