AI-Enhanced Turbidity Prediction: Mitigating Impacts on Dams Reservoirs and Water Treatment Plants

Sewoong Chung[1,*], Sungjin Kim[1], Dongmin Kim[1], Eunju Lee[1], Yeojeong Yun[1]

[1] *Dept of Environmental Engineering, Chungbuk National University, South Korea*

**Corresponding author:* schung@chungbuk.ac.kr

Abstract:

Climate change has led to an increase in extreme rainfall events in South Korea, resulting in intensified soil erosion and prolonged turbidity in dam reservoirs. This study aims to develop and evaluate an advanced turbidity prediction system that integrates a process-based water quality model (CE-QUAL-W2) with artificial intelligence and machine learning (AI/ML) models to address the limitations of traditional physical models. Focusing on the North Han River basin (South Korea), we constructed CE-QUAL-W2 model to generate training data for a Process Guided Deep Learning (PGDL) model. The PGDL approach enhances prediction accuracy while adhering to physical principles. In addition, we developed and compared various machine learning models for predicting inflow turbidity in dam reservoirs and water treatment plants. An integrated extreme rainfall-watershed runoff-turbidity prediction model is established, incorporating realistic scenarios based on climate change-induced rainfall pattern alterations. This model generates short-term runoff and turbidity prediction data. Additionally, we developed a user-friendly prototype program for turbidity prediction, featuring an intuitive interface and visualization tools for practical application by water resource managers. The resulting models and program serve as scientific tools for predicting and responding to turbidity events caused by extreme rainfall. This research contributes to improving water resource management efficiency, maintaining aquatic ecosystem health, and enhancing adaptability to climate change-induced alterations in the water environment.

Keywords: Extreme rainfall, Turbidity prediction, Process Guided Deep Learning, Climate change adaptation, Water resource management

Summary statement

This study develops an advanced turbidity prediction model integrating physical and AI/ML approaches for extreme rainfall events in South Korea, aiming to enhance climate change adaptation in the face of increasing environmental challenges.

1. Introduction

Recent increases in extreme rainfall events, driven by climate change, have intensified soil erosion in watersheds and river systems, thereby elevating suspended sediment concentrations and turbidity in rivers and reservoirs (Mimikou et al., 2000; Neff et al., 2000; Bouraoui et al., 2004). During such events, turbid water typically carries a diverse range of materials—including eroded and transported soil, resuspended sediments from riverbeds, and attached algae—underscoring the need for advanced modeling tools that link rainfall, watershed runoff, and turbidity transport.

Concurrently, rapid advancements in Fourth Industrial Revolution (4IR) technologies—most notably artificial intelligence (AI), machine learning (ML), big data analytics, and unmanned aerial systems—present new opportunities for accurate and timely turbidity monitoring and forecasting. Among data-driven models (DDMs), deep learning has demonstrated exceptional capacity for short-term predictions by leveraging extensive datasets; however, its inherent "black-box" nature can hinder the incorporation of fundamental conservation laws (e.g., mass, momentum, and energy). To address these shortcomings, an emerging strategy—referred to as "theory-based data science"—seeks to integrate process-based models (PBMs) with DDMs. Prominent applications include predicting lake water temperature (Read et al., 2019) and phosphorus levels (Hanson et al., 2020), highlighting the efficacy of combining mechanistic and data-driven approaches.

Although PBMs accurately simulate the spatiotemporal dynamics of turbidity by solving governing equations, they are often hampered by lengthy computation times and high sensitivity to initial and boundary conditions. To bridge these gaps and enhance predictive reliability, a hybrid turbidity forecasting framework that unites PBM insights with AI/ML techniques is indispensable. This study employs such an integrative approach to refine turbidity prediction and formulate a robust management strategy, particularly under intensifying hydroclimatic fluctuations.

2. Materials and Methods

This study was conducted in the Soyang Dam watershed and the North Han River basin in South Korea (Fig. 1), where turbidity problems due to extreme rainfall events continuously occur. The mechanical model and data-based model used for PGDL model development are CE-QUAL-W2 and LSTM (Long Short-Term Memory), a recurrent neural network deep learning model, respectively. Each model was calibrated and trained using temperature and turbidity data measured at the Paldang Dam and Soyang Dam from January to December 2020. Water temperature and turbidity data were collected from daily measurements provided by the Water Environment Information System of the Ministry of Environment and real-time automatic monitoring networks operated by K-water (Korea Water Resources Corporation). Real-time automatic monitoring data were collected hourly and converted to daily average values for model application. The main algorithm of the PGDL model for temperature and turbidity prediction optimized parameters by adding energy and mass balance terms from the mechanical model as constraints to the objective function (or loss function) of the LSTM model, imposing penalties when prediction results did not satisfy physical conservation laws. For the machine learning model to improve the boundary condition turbidity data of the CE-QUAL-W2 model, the LSTM model was used, with the Stepwise Multiple Linear Regression (SMLR) model as a comparison model. These models were trained using turbidity, flow rate, and precipitation data collected from the study area. To enable fair performance comparison between models, the period from January 1, 2016, to December 31, 2019, was set as the training period, and the period from January 1, 2020, to December 31, 2020, was set as the validation period for all models.

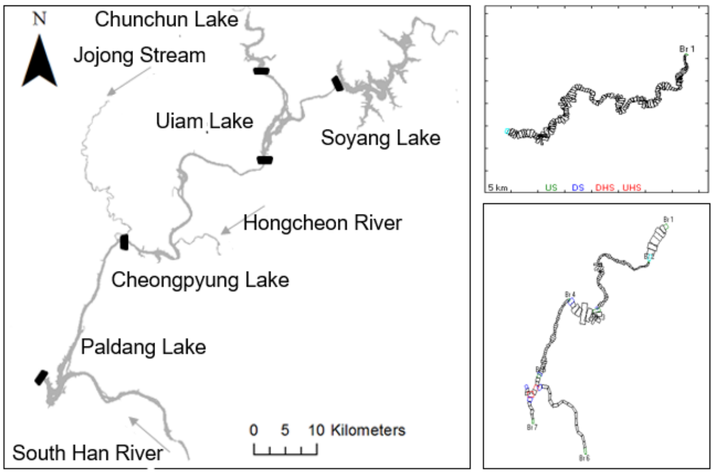

Figure 1 Study area map of the North Han River basin including Soyang Dam watershed in South Korea

Additionally, to predict turbidity at water treatment plants that intake water from the Paldang Lake, AI/ML models including Random Forest (RF), Support Vector Machine (SVM), XGBoost, Multiple Linear Regression (MLR), and Bidirectional LSTM (Bi-LSTM) were developed and evaluated using the same datasets.

3. Results and Discussion

3.1 Turbidity prediction using a hybrid process-based CE-QUAL-W2 and AI/ML approach

Comparative analysis of the LSTM model against the linear regression-based SMLR model yielded significant findings. Through detailed analysis of various error indicators including RMSE (Root Mean Square Error), NSE (Nash-Sutcliffe Efficiency), and adjusted coefficient of determination (Adj. R^2) during both training and validation phases, the LSTM model outperformed the SMLR model in predicting turbidity within reservoirs. On average, RMSE was 1.34 times lower, while NSE and Adj.R^2 improved by 4.36 and 1.28 times, respectively. We evaluated the LSTM model's applicability as a boundary condition turbidity prediction model by integrating it with the W2 model, using W2 model results with SMLR model turbidity predictions as a comparison. Results demonstrated that using the LSTM model as a boundary condition prediction model outperformed the SMLR model in reservoir turbidity prediction, with RMSE and RMSEN being 4.07 and 1.73 times lower on average, while model efficiency (measured by NSE values) improved by 1.46 times (Fig. 2).

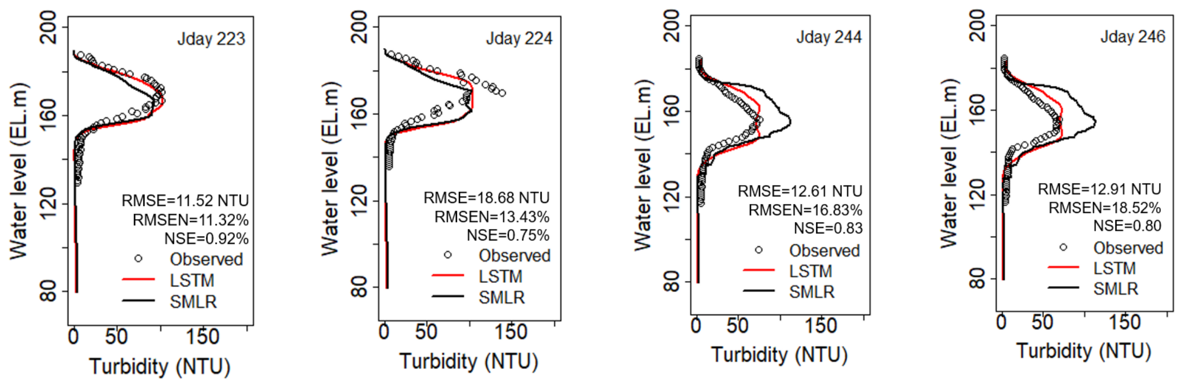

Figure 2 Comparison of turbidity profiles in Soyang Lake: LSTM vs. SMLR simulations

3.2 Turbidity prediction using a process-guided deep learning model

To evaluate the turbidity prediction performance of the PGDL model in Soyang Lake, we compared observed turbidity, PGDL model predictions (red line), and W2 model simulations (black line) together (Fig. 3). The PGDL model appropriately reproduced the measured turbidity values at different depths during both training and validation periods. The errors between predicted and observed values in PGDL model training and validation data were RMSE = 0.3 ~ 22.1 NTU, RMSEN = 10.5 ~ 34.8%, and RMSE = 0.4 ~ 23.7 NTU, RMSEN = 11.6 ~ 36.4%, respectively, demonstrating superior prediction performance compared to the W2 model (training data: RMSE = 5.0 ~ 22.8 NTU, RMSEN = 9.5 ~ 38.4%; validation data: RMSE = 5.3 ~ 23.9 NTU, RMSEN = 9.6 ~ 39.3%).

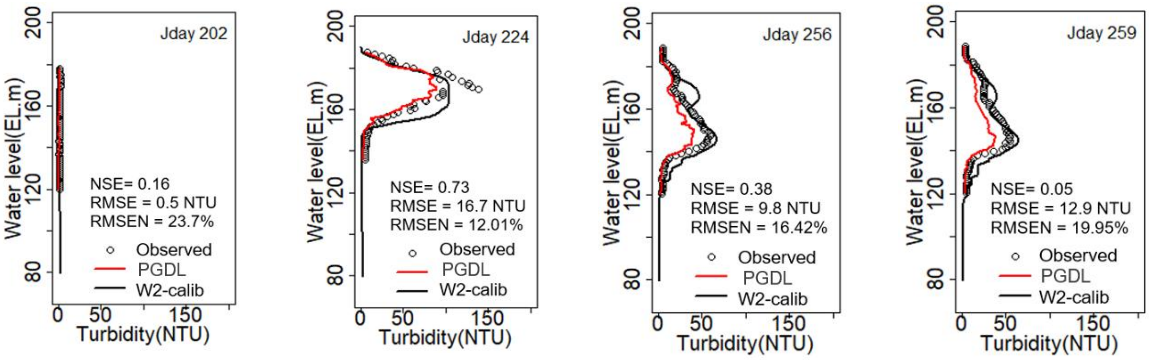

Figure 3 Depth-specific turbidity comparison among observed data, PGDL model predictions (red line), and W2 model simulations (black line) in Soyang Lake

3.3 Turbidity prediction in water treatment plants using AI/ML models

In predicting turbidity one day in advance (t+1), the LSTM and RF models achieved the highest performance, both with an R^2 of 0.77, closely followed by XGBoost (R^2 = 0.76) and Bi-LSTM (R^2 = 0.74). For two-day predictions (t+2), the LSTM model showed the highest accuracy (R^2 = 0.70), with XGBoost (R^2 = 0.69) and Bi-LSTM (R^2 = 0.64) performing similarly or slightly lower. Additionally, RF demonstrated the lowest error (RMSE = 6.5 NTU) for the one-day lead prediction, while LSTM exhibited the lowest error (RMSE = 7.8 NTU) for the two-day forecast. Notably, the MLR, XGBoost, RF, and SVM models showed a tendency to overfit, indicated by higher errors on the validation set compared to the training set, with prediction errors increasing as the forecast lead time extended.

Table 1 Performance of AI/ML models in predicting water treatment plant turbidity by lead time

Model	Lead time (day)	Train		Validation		Test	
		RMSE (NTU)	R^2	RMSE (NTU)	R^2	RMSE (NTU)	R^2
MLR	t+1	6.4	0.83	18.6	0.41	7.5	0.69
	t+2	10.1	0.61	17.9	0.30	12.7	0.39
XgBoost	t+1	3.8	0.94	8.6	0.73	7.4	0.76
	t+2	7.8	0.76	8.8	0.69	8.8	0.69
RF	t+1	3.1	0.96	17.9	0.46	6.5	0.77
	t+2	4.2	0.93	17.5	0.32	10.7	0.57
SVM	t+1	4.4	0.92	18.0	0.45	8.3	0.62
	t+2	5.5	0.86	15.7	0.45	12.4	0.42
LSTM	t+1	8.0	0.71	8.3	0.69	8.3	0.77
	t+2	9.0	0.64	9.3	0.62	7.8	0.70
Bi-LSTM	t+1	8.6	0.65	8.9	0.63	7.7	0.74
	t+2	9.3	0.62	9.5	0.61	9.4	0.64

3.4 A user-friendly, integrated turbidity prediction system

The integrated system combines three independently developed models—Precipitation Model, Watershed Runoff Model, and Turbidity Prediction Model in Rivers and Reservoirs—into a user-friendly, graphical user interface (GUI)-based application. These models operate sequentially, with the outputs from each preceding model serving as inputs for the next. In addition, the GUI incorporates external datasets required for model inputs, which are directly retrieved from officially published sources via Application Programming Interface (API) integration. This structure ensures seamless accessibility, practical usability, and intuitive operation for researchers and water resource management practitioners.

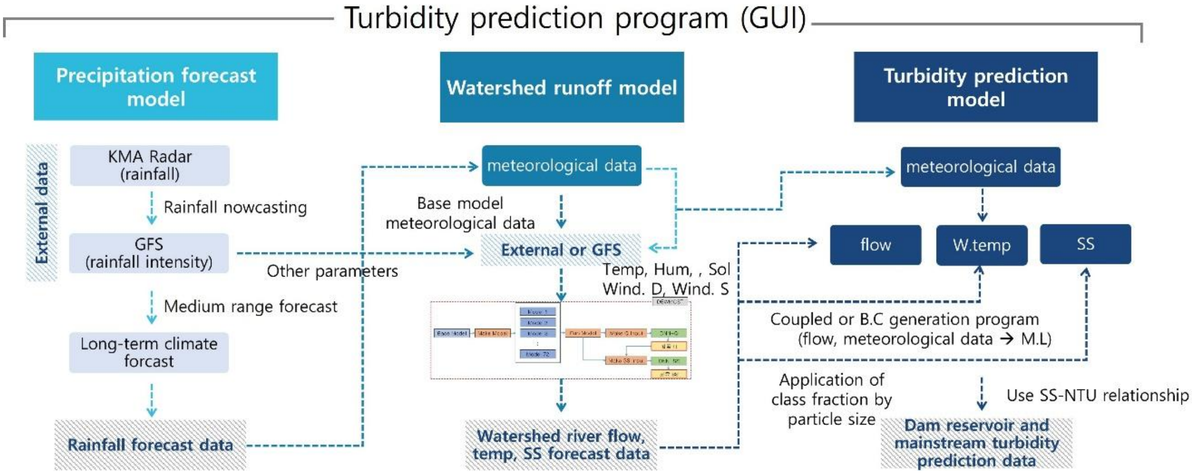

Figure 4 Inter-model coupling diagram in the turbidity prediction program (GUI)

4. Conclusions

This study developed and validated an integrated turbidity prediction approach combining process-based modeling (CE-QUAL-W2) and advanced AI/ML methods (LSTM, PGDL), significantly enhancing prediction accuracy compared to traditional models. The hybrid models showed superior performance in simulating turbidity at various depths in reservoirs and predicting inflow turbidity at water treatment plants. The developed user-friendly prediction platform facilitates practical turbidity management during extreme rainfall events, supporting effective water resource management and enhancing resilience against climate-induced water quality issues.

Acknowledgments: This work was supported by the Korea Environmental Industry & Technology Institute (KEITI) through the Aquatic Ecosystem Conservation Research Program, funded by Korea Ministry of Environment (MOE) (Grant number 2021003030004).

References

Bouraoui, F., Grizzetti, B., Granlund, K., Rekolainen, S., and Bidoglio, G. (2004). Impact of climate change on the water cycle and nutrient losses in a Finnish catchment, Climatic Change, 66(1), 109-126.

Mimikou, M. A., Baltas, E., Varanou, E., and Pantazis, K. (2000). Regional impacts of climate change on water resources quantity and quality indicators. Journal of Hydrology, 234(1-2), 95-109.

Neff, R., Chang, H., Knight, C., Najjar, R., Yarnal, B., and Walker, H. (2000). Impact of climate variation and change on Mid-Atlantic region hydrology and water resources, Climate Research, 14(3), 207-218.

Hanson, P.C.; Stillman, A.B.; Jia, X.; Karpatne, A.; Dugan, H.A.; Carey, C.C.; Stachelek, J.; Ward, N.K.; Zhang, Y.; Read, J.S.; et al. Predicting lake surface water phosphorus dynamics using process-guided machine learning. Ecol. Modell. 2020, 430, 109136.

Read, J.S.; Jia, X.; Willard, J.; Appling, A.P.; Zwart, J.A.; Oliver, S.K.; Karpatne, A.; Hansen, G.J.A.; Hanson, P.C.; Watkins, W.; et al. Process-guided deep learning predictions of Lake water temperature. Water Resour. Res. 2019, 55, 9173–9190.