Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

RIVER DIGITAL TWIN FOR WATER QUALITY PREDICTION

Surabhi Shrivastava[1], Souvik Barat[1], Shankar Kausley[1], Vinay Kulkarni[1], and Beena Rai[1]

[1]Tata Consultancy Service Research, Tata Consultancy Services, Pune, Maharashtra, INDIA

ABSTRACT

River is complex system of system, whose dynamics are influenced by multiple factors such as river characteristics (e.g., gradient and terrain), environmental factors (e.g., rainfall and temperature), and human interventions like building dams, and discharging wastewater. To understand and improve river water quality, a river digital twin is created using a combination of agent-based and physics-based models. Agent-based modeling captures behavioral relationships for ease in modeling complex systems and physics-based models simulate transport and reaction behavior. This integrated approach constructs a digital twin capable of simulating quality parameters under various scenarios, like rainfall, effluent discharge, changing demographics, and climate. The study enhances understanding of river ecosystems and provides a tool for managing their ecological health. The river digital twin is developed considering river and its ecosystem with different inflows and outflows and is applied successfully using a 480 km stretch of an India's largest river, the Ganga.

1 INTRODUCTION

Rivers serve as vital water sources for household, industrial, and irrigation needs worldwide, and support diverse ecosystems. The quality of river water impacts human well-being and the environment (Interlandi and Crockett 2003). Maintaining river water quality is essential, not just at specific points but along its entire course, for better human health across cities and villages, and the sustenance of the aquatic ecosystem (Anh et al. 2023). Despite good initial quality at the source, rivers often degrade due to industrial and municipal discharges, runoff, erosion, and various disruptions caused by human and environmental factors. These factors continuously shape the quality of the river water as it flows from its source to the river mouth (Ukhan and Luzovitska 2021). With proper planning and measures and leveraging the river's self-cleaning ability (Wen et al. 2017), river water quality can be preserved throughout its course. However, a comprehensive study of water quality related factors and understanding their influencing dynamics are crucial for accurately assessing, maintaining, and improving river water quality.

The current methods for assessing river water quality involve monitoring pH, conductivity, total dissolved solids (TDS), dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), and turbidity at spaced locations manually or with remote sensing (Demetillo et al. 2019). While government bodies and NGOs may also use automated techniques, these sporadic measurements don't offer a comprehensive understanding of the dynamics of water quality variations along the river (Huang et al. 2021; Gotovtsey 2010). Point measurements provide data on specific quality indicators, such as BOD and COD, but they lack insight into underlying factors explaining why the situation is so, and how it can be improved. Therefore, they often fail to produce meaningful justifications for deciding on effective interventions (Sutadian et al. 2016; Ewaid et al.2018).

Understanding the dynamics of water quality of river is complex. It requires a comprehensive view of river ecosystems along three dimensions: hydraulic flows and morphological characteristics, atmospheric environment of the river, and human-induced disruptions, along with their complex and nonlinear interferences. For example, river characteristics, such as flow velocity, inflow and outflow volumes, gradients, and the heterogeneous nature of riverbeds, play a significant role in influencing water quality.

979-8-3315-3420-2/24/$31.00 ©2024

Environmental factors, including weather conditions such as ambient temperature, rainfall, and humidity also contribute significantly. For instance, high temperatures can increase evaporation rates and alter physicochemical parameters, while intense rainfall can lead to runoff and pollutant transport into the river. Human-made interventions, such as discharge of polluted water from household, industrial, and agricultural activities, as well as flow disruptions caused by dams and diversions, further impact water quality. These multidimensional factors interact nonlinearly, changing over time and location, influencing neighboring river segments differently as the river flows. Conventional tools and techniques (Demetillo et al. 2019) struggle to capture these multi-dimensional influencing factors, heterogeneity, and spatiotemporal dynamics in a comprehensive manner. Current methodologies concentrate on individual or a limited set of characteristics without accounting for the interplay of multiple factors from both neutralization and aggregation viewpoints, as well as the heterogeneity within localized and neighboring contexts.

We present a simulatable digital twin, based to comprehensively represent and analyze all relevant factors related to the river, its environment, and possible human-made disruptions. We adopt an extended form of agent abstraction for capturing river as set of unique segments. The core of the digital twin is our simulatable agent model where each segment is represented as an agent having its own characteristics, behaviors, and interactions (with its neighboring segments), effectively capturing complex behaviors and heterogeneity. Instead of treating the river as a single agent, we consider it as a collection of agents interacting with each other in accordance with the physics governing the river flow to capture area-specific heterogeneity in a comprehensive way. We seamlessly integrate agent modelling paradigm with physics-based model within a simulatable framework to predict river quality accurately. In this formalism, the agent abstraction helps capture the inherent heterogeneity of various entities associated with the river ecosystem, while physics-based models are used to capture the natural phenomena of hydrology and water quality dynamics using mass-balance equations (Chen et al 2023).

The remainder of the paper is structured as follows: Section 2 provides an overview of state-of-the-art quality assessment techniques. In Section 3, we outline our approach. Section 4 showcases the application of our approach to an Indian river. Finally, Section 5 concludes the paper by discussing our early experiences, limitations, and avenues for future research.

2 LITERATURE REVIEW

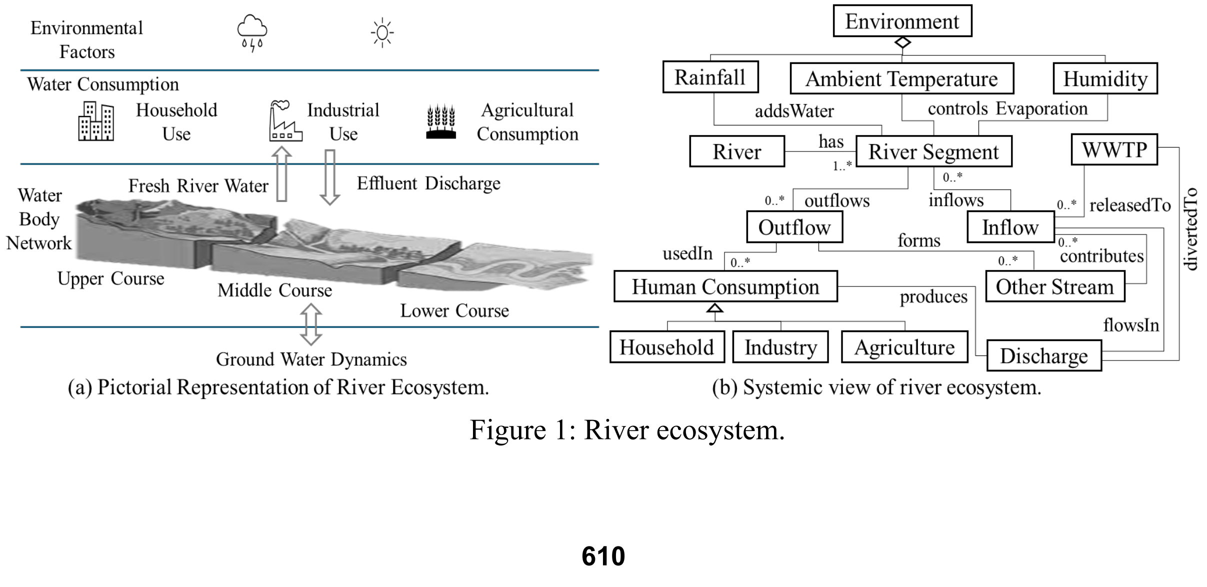
Various mathematical models are available in the literature for simulating and predicting the behavior of river water quality based on historical data. Methods such as the Mann-Kendall (MK) trend test (Kisi and Ay 2014), data-based models like Artificial Neural Network (ANN) (Abyaneh 2014; Singh et al. 2009), recurrent neural networks (RNN) (Li et al. 2019), and genetic programming (GP) (Jafari et al. 2020), support vector machine (SVM) (Leong et al. 2021), and neuro-fuzzy (NF) (Azad et al. 2019) models, have been developed as standalone approaches. Additionally, hybrid models have been created for water quality modeling and monitoring (Kisi and Parmar 2016; Abba et al. 2017). These models utilize extensive historical data to forecast the quality of the river at specific locations in the future. However, these methods have limitations. They cannot effectively model or predict phenomena that have not been observed in the past, rendering them inadequate for capturing evolving phenomena of the river ecosystem and their effects on water quality. Furthermore, data-based models are limited in their capability of finding the root cause of the factors behind deviations in river water quality (Asadollah et al. 2021). Models often point out only the key factors responsible for water quality which have direct correlation in the dataset. Several software options are also available for river modeling and predicting changes in river water quality.

However, these software programs primarily focus on hydrodynamics, flow patterns, and the influence of various factors on river flow and dispersion (Uddin et al. 2023; Ji 2017). These models are often rigid in nature and rely on pre-existing relationships and available information, lacking the flexibility to incorporate new parameters and effects. Software such as MIKE21 (Warren and Bach 1992) and CORIWAQ (Nguyen et al. 2018) are commonly used for river water modeling, but they primarily emphasize steady-state flows, flow dynamic behaviors, and water planning systems. The modeling of water quality parameters and their behavior is not given significant importance in these models. Furthermore, these models typically employ fixed quality parameters like biochemical oxygen demand (BOD), dissolved oxygen (DO) or chemical oxygen demand (COD), that cannot be adjusted based on the user's specific requirements for key parameter indices like total suspended solids (TSS), total dissolved solids (TDS) or turbidity.

3 APPROACH

The development of a river digital twin involves creating a simulated representation of the actual river, its ecosystem and environment, along with the human-made disruptions, closely resembling the real system as shown in Figure 1 (a). We consider a river ecosystem as a complex interplay among multiple natural and human induced elements across three dimensions: a) hydrological network, encompassing the river, its segments, and associated morphological features, b) environmental context, and c) water management dynamics extended for different consumptions. The key components and their interactions within a river ecosystem are depicted in Figure 1 (b). As shown, a river comprises multiple segments and various types of inflows and outflows that affect its flow dynamics and water quality. Chiefly, the river water flows outward to diverse entities, creating natural water streams and meeting various human needs, such as household, agricultural, and industrial purposes. Conversely, the river receives inflows from different natural streams, including tributaries at various confluences, as well as discharge from various water usage activities. These inflows of discharge can occur directly in an uncontrolled manner or be routed through wastewater treatment plants (WWTPs). Environmental factors such as rainfall, ambient temperature, and humidity influence water volume and quality through direct precipitation, runoff, and evaporation process.

Morphologically, a river begins its journey at the source and traverses a distance over heterogeneous terrains, experiencing various weather conditions, before it terminates at the river mouth. Typically, at the source, the narrowness and shallow depth of the riverbed contribute to swift flow. Mountainous regions often feature rocky riverbeds, resulting in more turbulent and energetic flows with higher reaeration capacity. As the river progresses and widens, the depth of the riverbed deepens, allowing for greater water volume and energy accumulation. Tributaries contribute to the river's flow along its journey, merging at confluences, which alter flow patterns and sediment loads. In the middle course, the composition of the riverbed influences erosion and deposition patterns, with softer sedimentary riverbeds prone to erosion and the formation of meander bends. These characteristics also impact the formation of oxbow lakes, as wider and shallower sections encourage sediment deposition, leading to the formation of cutoff meander loops. In the lower course, the breadth of the riverbed widens significantly as it nears the sea or ocean, facilitating the formation of expansive deltas where sediment deposition occurs. In estuaries, the depth and composition of the riverbed plays a crucial role in determining the mixing of freshwater and saltwater, affecting salinity levels and ecosystem dynamics. Throughout its journey, the river may encounter several rapids, influenced by the depth, composition, and shape of the riverbed, with shallower sections and rocky substrates creating more turbulent, rapids, and different flow patterns affecting river water quality. Additionally, the river may



(a) Pictorial Representation of River Ecosystem.

(b) Systemic view of river ecosystem.

Figure 1: River ecosystem.

receive water through runoff from rain and flooding in neighboring areas, altering the depth and composition of the riverbed and affecting the severity of flooding and the distribution of sediment deposition.

We adopt a simulatable agent abstraction and a multi-modeling technique, involving agent modeling, data-centric models, and physics-based models synergistically, designed to comprehensively model the river ecosystem. We utilize various data sensors located at different parts of the river ecosystem to measure inflow and outflow volumes and quality, along with weather data, to synchronize the constructed model with the real system. The synchronized model of the river ecosystem forms the digital twin of the river ecosystem, which is used for simulation to experiment with potential strategies aimed at improving water quality, such as installing and positioning WWTPs and other forms of river cleaning techniques, such as sediment traps. The remaining section discusses our modeling approach, schema for data synchronization, and an implementation of a river ecosystem digital twin.

3.1 River Ecosystem Model

We utilize an extended version of the Agent/Actor model (Clark et al. 2017) to represent the active elements of river ecosystem as depicted in Figure 1 (b). The central element of a river ecosystem is the river itself, along with all streams that flow into and out of it. While in actual, a river is a unified entity where water flows from its source to the river mouth, we conceptualize it as a sequence of interconnected segments, each represented as an Agent. The key abstractions and agent definitions of the river ecosystem model are illustrated in Figure 2.

We define a river agent as comprising an ordered set of segment agents having a specific water quality measure defined in terms of BOD and DO. In a river, the quality and flow dynamics of a river segment influence its downstream segments. Incoming streams also affect the quality and flow dynamics, whereas outgoing streams from a segment only influence the flow dynamics. Each segment agent (and its quality) is characterized by a specific architype, defined based on a set of factors including - a) Geometric characteristics (e.g., length, width, depth, flow rate), b) Riverbed composition (e.g., Rocky, Sandy, Sedimentary), c) Various morphological features (as depicted in Figure 2), d) The number and types of inflowing and outflowing streams, and e) Environmental factors. We consider three categories of environmental factors: Precipitation and humidity (e.g., Wetland, Humid, Dry) and temperature profile of segment locality (e.g., Polar, Temperate, Tropical).

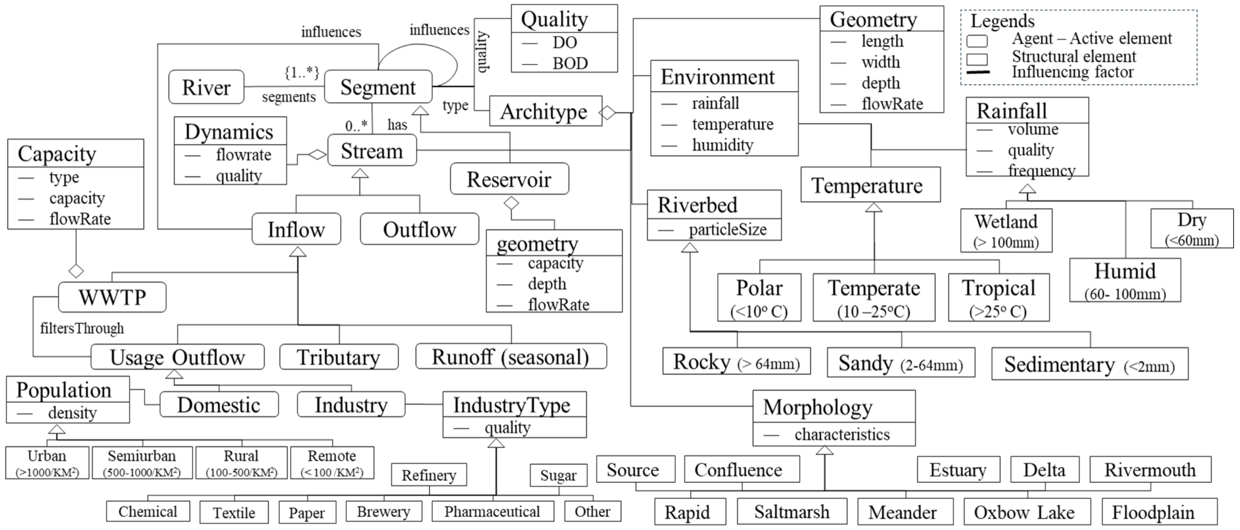
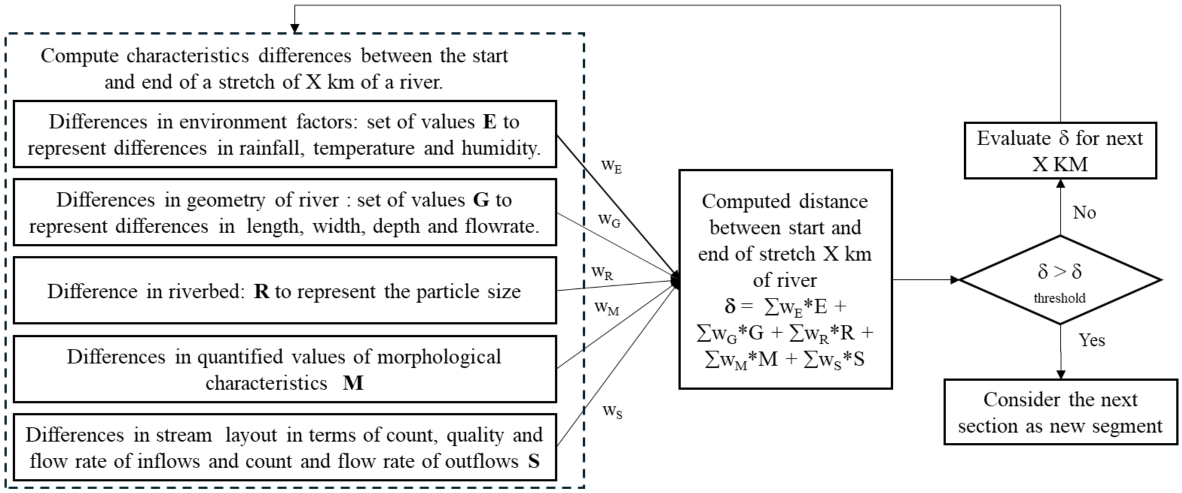


Figure 2: Modelling abstractions for river digital twin ecosystem.

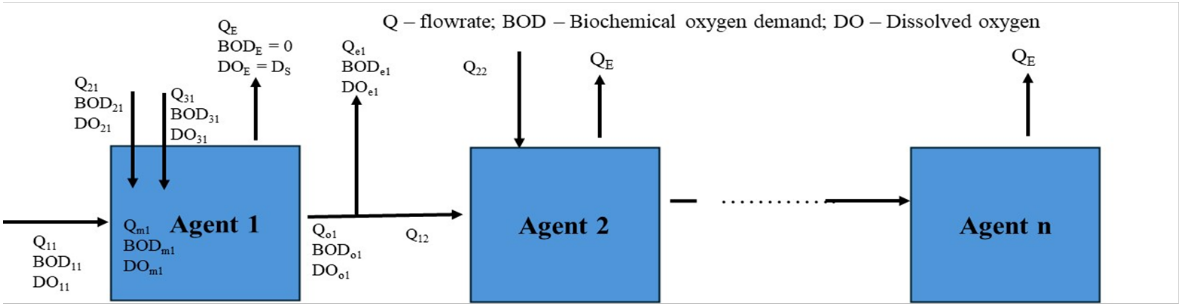
Inflow streams, which influence both water quality and flow dynamics, are typically categorized into four types: Tributaries joining the river, runoffs from local areas due to rainfall and flooding, direct outflows from different types of water consumption sources or usage and treated water outflows from WWTPs. Water usages are further classified, including domestic and industrial uses, and can be further categorized based on population density and industry types as depicted in Figure 2. These architypes are defined for industry and domestic usage patterns. For example, domestic architypes based on population density indicate directly influence wastewater quality and volume. Areas with high density and less compliance, e.g., suburban areas, may lack treatment facilities, resulting in highly polluted water discharge. Industry architypes are classified by industry type and treatment methods, affecting effluent discharge and environmental impact. Additional architypes can be added under existing or new headers to accommodate advanced systems. For example, usage outflow may also include agricultural usage.

3.1.1 River Model

In our model, we capture the heterogeneities along geometric, morphological, weather, and terrain and stream topology as shown in Figure 2. In reality, these characteristics may continuously change, resulting in infinitely many segments characterized by segment architypes. To manage this complexity, we employ an algorithm, as depicted in Figure 3 (a), to derive a manageable set of river segments. Conceptually, the number of river segments is determined by noticeable differences of characteristics along a particular stretch of river. We compute the difference (δ) between the weighted sums of all characteristic values at the beginning and end of the stretch of X km which is taken iteratively from the start of the river in



a) Algorithm to calculate number of river segments.



b) Illustration of agent behaviour and interactions.

Figure 3: Agent, agent segment and physical-based behaviour.

successive X multiples. The value of X depends on the river system and its heterogeneity. For the case of Ganga River simulations X is taken to be 10 km. The values at the start and end of the stretch come from the river environmental and characteristic data from the literature and database and the weights are the average value of the particular parameter in the whole stretch. We then compare this difference with a threshold value ($\delta_{threshold}$) to determine whether the next section will be considered a new segment agent or not. In the case of Ganga River, the $\delta_{threshold}$ is 5. We classify a new segment if δ value exceeds the threshold value and assign architype based on the average values of all associated characteristics. Consequently, the river model can be envisioned as a connected set of segments, each characterized by its respective architype.

In this formulation, different segments of the river interact with each other according to various sets of rules and interactions. For river interactions within itself and external agents, the rules are physics-based, and others are data-based and empirical rules taken from different literature and are discussed in Section 3.1.3.

3.1.2 Physics Based Equation for River Dynamics

As shown in Figure 2, each segment exhibits a specific quality parameter (e.g., BOD and DO) and the dynamics of these quality parameters are governed by the law of physics and hydrology. We define a set of mass balance equations to represent the dynamics of the quality parameters (BOD and DO). Primarily it is a weighted average of mixing streams (Equations (1)-(3)), ensuring precise representation of mixing effects on water quality.

$$Q_{mi} = Q_{1i} + Q_{2i} + Q_{3i} + \cdots Q_{li} - Q_E \tag{1}$$

Equation (1) represents the mixing of different streams at the start of an i^{th} river agent, where Q1i is an inlet from the previous agent of river/reservoir and Q_{2i} to Q_{li} is inlet from external agents (industries, WWTP, rainfall), Q_E is volume loss by evaporation. Quality parameters i.e., BOD and DO change due to mixing is represented by Equations (2) and (3).

$$BOD_{mi} = \frac{Q_{1i} \times BOD_{1\mathrm{i}} + Q_{2\mathrm{i}} \times BOD_{2\mathrm{i}} + Q_{3i} \times BOD_{3i} + \cdots Q_{li} \times BOD_{li}}{Q_{mi}} \tag{2}$$

$$DO_{mi} = \frac{Q_{1i} \times DO_{1\mathrm{i}} + Q_{2\mathrm{i}} \times DO_{2\mathrm{i}} + Q_{3i} \times DO_{3i} + \cdots Q_{li} \times DO_{li}}{Q_{mi}} \tag{3}$$

The evaporation from the river surface can be estimated using the modified Penman-Monteith equation, suggested by Valiantzas in his work (Valiantzas 2006). The evaporation is calculated in mm/day can be converted to m^3/day by multiplying evaporation by length (L) and width (W) of the river of that agent. The BOD and DO of the river agent is instantiated by the mass balance equation in (2) and (3), but it changes as it leaves the agent to the other agents according to Equations (4) and (5).

$$BOD_{oi} = \frac{BOD_{mi}}{(1 + k_{di}\tau_i)} \tag{4}$$

$$DO_{oi} = \frac{DO_{mi} + k_{ri}D_s\tau_i - k_{di}\left(\frac{BOD_{mi}}{1 + k_{di}\tau_i}\right)\tau_i}{(1 + k_{ri}\tau_i)} \tag{5}$$

where k_{di} and k_{ri} are the BOD degradation coefficient and reaeration coefficient of the river section/agent which depends on the temperature, velocity, riverbed properties. BOD_{mi} and DO_{mi} is the mixed BOD and DO values, respectively calculated from mass balance equations at the inlet of any river agent. Ds is the saturation DO, which is dependent on the temperature of the river section. τ_i is calculated by dividing volume of the river section (L*W*D) with the total flow in the river section Q_{mi}. The pictorial representation of the interaction of different agents is given in Figure 3 (b). The reaeration rate (k_r) is a function of the velocity and bed height of the river section and deoxygenation rate (k_d) is a function of temperature. Similarly, any other physics of the system can be added in the river class to include more dynamics or behavior in the river system such algal growth or erosion of the riverbed etc.

3.1.3 Correlation Between Agents

Effluents from industry, streams, household, and WWTP are added or taken up by the river. The reservoir may discharge or take up water based on location and factors like height. Rainfall adds water directly to the river, with its volume determined by Equation (6) (Kostic et al. 2016).

$$\ln\left(V_{rain}\right) = 4.856 + (0.035 \times P - 0.0769 \times T) + (1.137 \times 10^{-6} \times P \times T) - (2.296 \times 10^{-4} \times P^2) \\ + (1.098 \times 10^{-3} \times T^2) + (1.152 \times 10^{-5} \times P^2 \times T) - (5.931 \times 10^{-5} \times P \times T^2 \qquad (6)$$

where P is monthly average rainfall in mm and V_{rain} is water influx in river due to rainfall m^3/day. The above equation can be converted to daily rainfall data for finding daily influx of water in river due to rainfall.

In certain scenarios, the household and industry agents may also directly interact with the wastewater treatment plant (WWTP). Depending on the situation, they may discharge water to the WWTP, which subsequently releases treated water back into the river. The WWTP, in turn, receives input from industry or household agents. Additionally, the rainfall contributes to the water influx in the WWTP. The model accounts for the influx of water in the WWTP due to rainfall using equation (7) (Mesdaghinia et al. 2015). This equation enables the accurate representation of the impact of rainfall on the WWTP's water inflow.

$$V_{rainWWTP} = (1.48P + 935.35) \times 1000 \qquad (7)$$

3.2 Illustration

We illustrate relevant concepts and formulation of digital twin using a unidirectional river with a length of 48 km as shown in Figure 4 for easier understanding. The river digital twin development illustrated in this example can be extended to bigger river systems and multiple other factors. The river originates from a source and terminates at the mouth. The river is assumed to have a one-way flow without any backflow or intermixing. Two different industries are situated at distances of 28 km and 36 km from the starting point of the river. The ecosystem includes a rural area that collects and redirects all its wastewater to a WWTP, which subsequently releases the treated water into the river at a location 18 km from the starting point. Additionally, a stream originates from a non-point source located 23 km from the starting point of the river. Industry 1, a brewery situated at the 28 km mark, also withdraws water from the river. The schematic diagram of the representative river is given in Figure 3. The riverbed, morphology, and weather change as

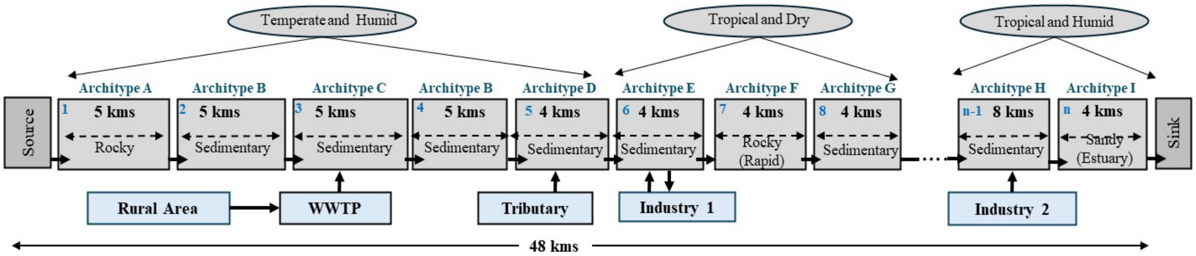


Figure 4: Illustration of a river ecosystem.

Table 1: Data required for modeling digital twin, its source and update frequency.

Data type	**Data source**	**Update frequency**
Temperature, humidity	Meteorological data	Hourly
Rainfall	Meteorological data	Daily
Flow volumes and quality	Sensory data and testing	Daily
River-bed characteristics	Reports	Annually
Population Demographic	Reports	Quarterly

it flows. The weather of the river's environment varies as follows- for the first at 24 km, it is temperate and humid, next 12 km is tropical and dry and final 12 km is tropical and humid. We represent this stretch of the river using 10 segments and 9 segment architypes as shown in Figure 3. In this formulation, Segment 1 (Architype A) and Segment 2 (Architype B) are different due to riverbed characteristics, Segment 2 and Segment 3 (Architype C) are different due to WWTP inflow dynamics, segments 2 and 4 are same architype, segment 5 differs from segment 2 and 3 due to tributary inflow. Segments 6, 7 and 8 are different from earlier segments as weather is different and they have different inflow characteristics. Here, Segments 1 and 7 are morphologically similar but belongs to different architypes due to weather. The weather of segment 9 and 10 differs from earlier. Segment 9 differs from segment 10 in terms of riverbed and morphology.

3.3 Digital Twin Construction and Data Synchronization

We employ a systematic process, as depicted in Figure 5, to construct a digital twin of a river ecosystem. Initially, we model the river or a section of it by utilizing the abstractions outlined in Figure 2. This entails replicating the river's geometry, riverbed, morphology, and topological features, including input and output streams along with their respective characteristics. Subsequently, we determine the required number of segment agents, formulate agents and define their architypes using the algorithm presented in Figure 3 (a). Real-world data, obtained from the system of records as highlighted in Table 1, is then integrated into the constructed model to develop an accurate digital twin of the river ecosystem. This data may originate from various sources and adhere to the schema highlighted in Figure 5. Data for the river properties such as riverbed characteristics, morphology, and geometry (length, width, and depth) come from different reports published by pollution control bodies, NGOs, and government agencies. These properties are not very dynamic in nature and often are constant for perennial rivers. For rain-fed rivers, the seasonal variations are high. Real-time values as well as historical data can be used depending on the river section, we are modelling. Different inflows and outflows for almost all rivers are monitored, and reports are published on the open sites with daily input/output volume, quality, and location of these effluent discharge. Data can be extracted from these reports for various inflows and outflows, their point of flow, quality and volumes

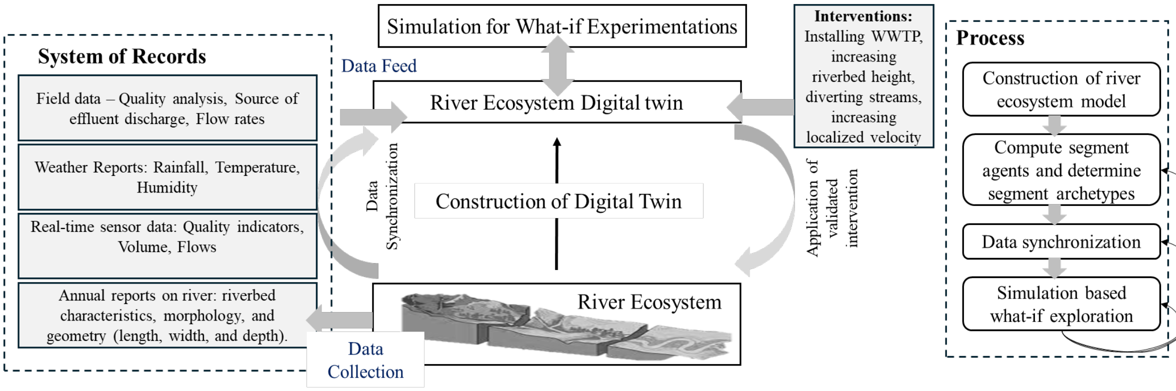


Figure 5: Construction, data synchronization and use of river ecosystem digital twin.

discharged. In addition, volume of effluent entering, their quality indicators can be loaded from the real-time data sensory data. At regular intervals of rivers, there are sampling points which collect regular quality data, analyse the deviation in quality and keep a check on the effluents getting added in the main streams. These field data is also well-documented and can be collected for modelling the river ecosystem. Environmental data such as temperature, relative humidity and rainfall are directly taken from the weather reports for each area and are updated in real-time. Data about type of household, population demographics and type of industries are also taken from published reports. Most of the published reports are freely available online and required data can be scraped from these sources.

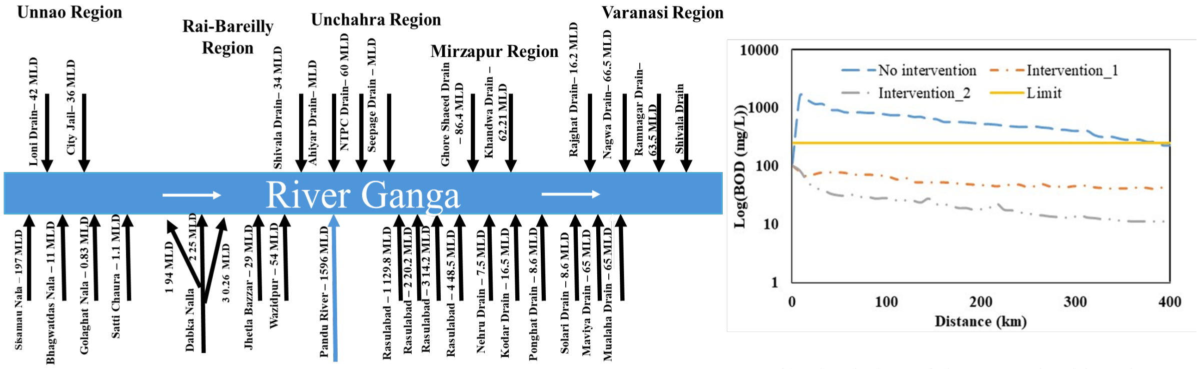
A fully functional digital twin synchronized with real data can then be simulated for what-if experimentations such as effect of different interventions, changing scenarios in case of new industries, population explosion and stringent guidelines by the pollution control bodies. Different interventions can be checked on the digital twin to identify the intervention best suited for the current situation. At the beginning of the simulation initial values of the river can be taken from the river water quality reports or a particular value can be assumed.

We simulate instantiated river model, i.e., river digital twin, to perform experimentations. The agent-based simulation is implemented using the MESA framework package available in Python (Masad 2015). Different agents are defined using classes, which allow for the encoding of their respective behaviors. The simulation progresses in discrete time steps, with each step equivalent to one day. To ensure this, all effluent discharge volumes, rainfall volumes, and other quantities are expressed in metric cubic meters per day.

4 APPLICATION

To assess the validity and applicability of our model, we utilized data from the Ganga River (Singh et al. 2007; Singh 2022) stretch of 480 km between Kanpur and Varanasi, out of its total length of 2525 km. The pollution assessment data was sourced from the Central Pollution Control Board (CPCB) document (CPCB 2013), which provides BOD values and daily effluent discharge volumes for 31 different streams that contribute to the river Ganga. Locations of these streams were obtained from the CPCB website. Only the major streams which are getting added in this stretch are considered. Small streams and non-point sources of pollution are not considered for this simulation. However, multiple non-point sources of pollution are being added in this stretch of water as the population residing on the banks use river water for recreational, utility, and spiritual activities. Most of the streams are the major 'nallas' or drainages of towns and cities of that area. The drains contain the wastewater of all the nearby municipal area as well as small industries. As can be seen from the data (Figure 6 (a)), the first drain, which is Sisamau Nalla has the maximum BOD load (product of BOD and volume discharge) followed by Wazidpur Nalla and Dabka Nalla. In a stretch of 480 km, approximately 30 streams with high BOD load are getting added in the main river at regular intervals. A major wastewater stream from NTPC plant is getting added 200 km downstream of the start of the river and high volume of water from Pandu River (tributary of River Ganga) and Yamuna River get added to the main river stream daily. Coal fields are situated on the Pandu River leading to generally high BOD levels of the river. This stretch is considered for simulating the digital twin since multiple streams are added at short intervals and the quality and quantity of the streams vary across a large range of values. This makes the estimation of quality parameters of the river stretch by conventional methods of sampling and testing difficult. Also, information and understanding of the streams responsible for major pollution and the necessary action which will result in overall better quality cannot be gathered without a detailed simulation of the stretch with different scenarios and interventions. Applicability of the model for such diverse conditions will validate the model and can be transferred to different rivers and ecosystem easily.

Accordingly, we created 30 stream agents to represent the various contributing streams (external agents) to the main river agent at different locations from Figure 6 (a). The 480 km stretch of Ganga from Kanpur to Varanasi is divided into different number of agents according to algorithm given in Figure 3 (a). The average values of environment data (temperature, humidity, and rainfall) were taken for the month of October. 31 number of segments were calculated from the algorithm and for 31 segments 9 different architypes were taken for the 480 km stretch. The simulation was conducted over a 100-day period. The



a) River Ganga ecosystem for Kanpur to Varanasi (CPCB 2013).

b) Simulation of river stretch with and without interventions.

Figure 6: Strech of Ganga from Kanpur to Varanasi.

different architypes differentiation is mainly on the basis on the number and quality of inflows in different sections since other parameters such as riverbed composition, environmental factors and morphology are almost similar in this stretch. The initial values of the BOD and DO of the stretch taken from the quality values published in reports for different sampling points.

Following the creation of digital twin for the stretch, the digital twin was simulated for a large period (100 days) to understand the long-term effect on the river water quality. The log of quality values of each section was plotted against the distance of the section from the start of the stretch. For comparison purposes, CPCB limit for the quality is also plotted on the same graph to check the sections which are crossing the set limit value of 50 mg/L.

As can be seen from Figure 6 (b), in the current condition of quality and quantity of streams adding, the overall quality of the river Ganga stretch is way above the specified limits. Hence, necessary actions and interventions need to be taken to improve the quality and keep it under check. To explore the effectiveness of interventions, two different interventions or actions were taken to limit the quality within the desired limits. For the first intervention, 100 mg/L BOD limit for all the streams was imposed. No stream can discharge water greater than 100 mg/L. This can be achieved by at source treatment or treatment before they enter the river stretch. The second intervention is when a limit of 50 mg/L was imposed for all incoming streams. The system was simulated with two different interventions to highlight the importance of changing norms and regulations and in making the right choice. Figure 6 (b) shows the variation of BOD in logarithmic scale with two interventions, without intervention. The second intervention showed better performance as the incoming BOD limit was decreased for all the incoming streams from intervention 1, whereas very high BOD values can be seen when there is no limit on the BOD value of the effluent stream discharged in the river. Other similar interventions like creating WWTP on the course of the river, increasing riverbed height, diverting streams to different locations can be simulated on the digital twin to choose the best possible intervention in a given scenario for a particular river stretch. Additionally, the digital twin can be simulated for different environmental conditions to understand its effects on the overall quality and helpful in deciding preventive actions to maintain the quality within the desired limits.

5 CONCLUSION

We discussed the importance of comprehensively understanding water quality dynamics within river ecosystems. To address this challenge, we proposed a simulatable model as the foundation for constructing a comprehensive digital twin of river ecosystems. By seamlessly integrating the agent modeling paradigm with physics-based and data-centric models within a simulatable framework, our approach enables precise prediction of river quality dynamics. We have also demonstrated the effectiveness of using agent abstraction to capture the inherent heterogeneity of entities associated with river ecosystems, coupled with seamless interoperability with physics-based models to accurately represent hydrological phenomena and water quality dynamics using mass-balance equations.

To validate our approach, we applied it to a 480 km stretch of the River Ganga, accurately simulating key parameters such as BOD and DO. The successful simulation of interventions on the River Ganga has highlighted the potential of our approach to support water quality management efforts. However, it is crucial to acknowledge certain limitations, including the ongoing need for calibration and validation to ensure the accuracy and reliability of the model, particularly concerning the definition of segments and segment architypes. Future research will explore enhancements to the model's predictive capabilities, the integration of real-time data sources for improved accuracy, and the consideration of socio-economic factors influencing river ecosystem dynamics.

REFERENCES

Abba, S. I., S. J. Hadi and J. Abdullahi. 2017. "River Water Modelling Prediction using Multi-linear Regression, Artificial Neural Network, and Adaptive Neuro-fuzzy Inference System Techniques". *Procedia Computer Science* 120: 75-82.

Abyaneh, H. Z. 2014. "Evaluation of Multivariate Linear Regression and Artificial Neural Networks in Prediction of Water Quality Parameters". *Journal of Environmental Health Science and Engineering* 12: 1-8.

Anh, N. T., N. T. Nhan, B. Schmalz and T. L. Luu. 2023. "Influences of Key Factors on River Water Quality in Urban and Rural Areas: A Review". *Case Studies in Chemical and Environmental Engineering* 100424.

Asadollah, S. B. H. S., A. Sharafati, D. Motta and Z. M. Yaseen. 2021. "River Water Quality Index Prediction and Uncertainty Analysis: A Comparative Study of Machine Learning Models". *Journal of Environmental Chemical Engineering* 9(1): 104599.

Azad, A., H. Karami, S. Farzin, S. F. Mousavi and O. Kisi. 2019. "Modeling River Water Quality Parameters using Modified Adaptive Neuro Fuzzy Inference System". *Water Science and Engineering* 12(1): 45-54.

Chen, S., J. Huang, P. Wang, X. Tang, and Z. Zhang. 2023. "A Coupled Model to Improve River Water Quality Prediction towards Addressing Non-stationarity and Data Limitation". *Water Research* 120895.

Clark, T., V. Kulkarni, S. Barat and B. Barn. 2017. "ESL: An Actor-based Platform for Developing Emergent Behaviour Organisation Simulations". *Advances in Practical Applications of Cyber-Physical Multi-Agent Systems: The PAAMS Collection: 15th International Conference, PAAMS 2017, Porto, Portugal, June 21-23, 2017, Proceedings 15*: 311-315.

Central Pollution Control Board. 2013. "Pollution Assessment of Ganga". https://cpcb.nic.in/wqm/pollution-assessment-ganga-2013.pdf Accessed on 18th April 2024.

Demetillo, A. T., M. V. Japitana and E. B. Taboada. 2019. "A System for Monitoring Water Quality in a Large Aquatic Area using Wireless Sensor Network Technology". *Sustainable Environment Research* 29: 1-9.

Ewaid, S. H., S. A. Abed and S. A. Kadhum. 2018. "Predicting the Tigris River Water Quality within Baghdad, Iraq by using Water Quality Index and Regression Analysis". *Environmental Technology & Innovation* 11: 390-398.

Gotovtsev, A. V. 2010. "Modification of the Streeter-Phelps System with the Aim to Account for the Feedback between Dissolved Oxygen Concentration and Organic Matter Oxidation Rate". *Water Resources* 37: 245-251.

Huang, J., Y. Zhang, H. Bing, J. Peng, F. Dong, J. Gao *et al.* 2021. "Characterizing the River Water Quality in China: Recent Progress and On-going Challenges". *Water Research* 201: 117309.

Interlandi, S. J. and C. S. Crockett. 2003. "Recent Water Quality Trends in the Schuylkill River, Pennsylvania, USA: a Preliminary Assessment of the Relative Influences of Climate, River Discharge and Suburban Development". *Water Research* 37(8): 1737-1748.

Jafari, H., T. Rajaee and O. Kisi. 2020. "Improved Water Quality Prediction with Hybrid Wavelet-Genetic Programming Model and Shannon Entropy". *Natural Resources Research* 29: 3819-3840.

Ji, Z. G. 2017. *Hydrodynamics and Water Quality: Modeling Rivers, Lakes, and Estuaries*. John Wiley & Sons.

Kisi, O. and M. Ay. 2014. "Comparison of Mann–Kendall and Innovative Trend Method for Water Quality Parameters of The Kizilirmak River, Turkey". *Journal of Hydrology* 513: 362-375.

Kisi, O. and K. S. Parmar. 2016. "Application of Least Square Support Vector Machine and Multivariate Adaptive Regression Spline Models in Long Term Prediction of River Water Pollution". *Journal of Hydrology* 534: 104-112.

Kostic, S., M. Stojkovic, S. Prohaska, and N. Vasovic. 2016. "Modeling of River Flow Rate as a Function of Rainfall and Temperature using Response Surface Methodology Based on Historical Time Series". *Journal of Hydroinformatics* 18(4): 651-665.

Leong, W. C., A. Bahadori, J. Zhang, and Z. Ahmad. 2021. "Prediction of Water Quality Index (WQI) using Support Vector Machine (SVM) and Least Square-Support Vector Machine (LS-SVM)". *International Journal of River Basin Management* 19(2): 149-156.

Li, L., P. Jiang, H. Xu, G. Lin, D. Guo and H. Wu. 2019. "Water Quality Prediction Based on Recurrent Neural Network and Improved Evidence Theory: A Case Study of Qiantang River, China". *Environmental Science and Pollution Research* 26: 19879-19896.

Masad, D. and J. L. Kazil. 2015. "Mesa: An Agent-Based Modeling Framework". *SciPy*: 51-58.

Mesdaghinia, A., S. Nasseri, A. H. Mahvi, H. R. Tashauoei and M. Hadi. 2015. "The Estimation of Per Capita Loadings of Domestic Wastewater in Tehran". *Journal of Environmental Health Science and Engineering* 13: 1-9.

Nguyen, T. T., I. Keupers and P. Willems. 2018. "Conceptual River Water Quality Model with Flexible Model Structure". *Environmental Modelling & Software* 104: 102-117.

Singh, K. P., A. Basant, A. Malik and G. Jain. 2009. "Artificial Neural Network Modeling of The River Water Quality—A Case Study". *Ecological Modelling* 220(6): 888-895

Singh, M., I. B. Singh and G. Müller, G. 2007. "Sediment Characteristics and Transportation Dynamics of The Ganga River". *Geomorphology* 86(1-2): 144-175.

Singh, I. B. 2022. "The Ganga River". *Large Rivers: Geomorphology and Management, Second Edition*: 521-550.

Sutadian, A. D., N. Muttil, A. G. Yilmaz and B. J. C. Perera. 2016. "Development of River Water Quality Indices—A Review". *Environmental Monitoring and Assessment* 188: 1-29.

Uddin, M. G., S. Nash, A. Rahman and A. I. Olbert. 2023. "Performance Analysis of the Water Quality Index Model for Predicting Water State using Machine Learning Techniques". *Process Safety and Environmental Protection* 169: 808-828.

Ukhan, O. O. and Y. A. Luzovitska. 2021. "Integrated Assessment of the Southern Bug River Basin Water Quality and Calculation of their Self-Cleaning Capacity". *Man and Environment. Issues of Neoecology* 35: 48-57.

Valiantzas, J. D. 2006. "Simplified Versions for the Penman Evaporation Equation using Routine Weather Data". *Journal of Hydrology* 331(3-4): 690-702.

Warren, I. R. and H. Bach. 1992. "MIKE 21: A Modelling System for Estuaries, Coastal Waters and Seas". *Environmental Software* 7(4): 229-240.

Wen, Y., G. Schoups and N. Van De Giesen. 2017. "Organic Pollution of Rivers: Combined Threats of Urbanization, Livestock Farming and Global Climate Change". *Scientific Reports* 7(1): 43289.

AUTHOR BIOGRAPHIES

SURABHI SHRIVASTAVA is a Researcher at Tata Consultancy Services (TCS) Research, Pune, India. Her current research interests include modelling, simulation, optimization, and analysis of engineering processes. Her email address is shrivastava.surabhi2@tcs.com.

SOUVIK BARAT is a Principal Scientist at Tata Consultancy Services (TCS) Research, India, visiting researcher at Middlesex University London and adjunct professor at Walchand College of Engineering. His research interests include Digital Twin technology, Modelling and simulation, and Model Driven Engineering. His email address is souvik.barat@tcs.com.

SHANKAR B. KAUSLEY is a Senior Scientist at Tata Consultancy Services (TCS) Research, India. His current research interests include water and wastewater treatment, food quality prediction and functional coatings. He is a 'Certified Water Specialist' from Water quality association, USA and is a member of the Global Young Academy (GYA), Germany. His email address is shankar.kausley@tcs.com.

VINAY KULKARNI is a TCS Fellow at TCS Research, Pune, India. His current research interests are Digital Twins, Adaptive Enterprises, and Software Engineering. Alumnus of Indian Institute of Technology Madras, Vinay is a Fellow of Indian Academy of Engineering. His email address is vinay.vkulkarni@tcs.com.

BEENA RAI is Chief Scientist at TCS Research. Her research interests include Computational Materials Science, Materials Informatics and Digital Twins & Threads. Beena is a Fellow of Indian Academy of Engineering. Her email address is beena.rai@tcs.com.