Using Cyber Physical Systems To Map Water Quality Over Large Water Bodies

P. Hirani¹, S. Balivada¹, R. Chauhan¹, G. Shaikh¹, L. Murthy¹, A. Balhara¹, R. C. Ponduru¹, H. Sharma¹, S. Chary², G. B. Subramanyam², S. Randhawa³, T. Dutta³, H. P. Gupta⁴, A. Gupta⁵, A. Haldar⁵, A. Sarkar⁵, I. Khan⁵, S. Guha⁶

¹ University of Chicago Center in Delhi, New Delhi, India
² IBM Research, Bangalore, India
³ Dept. of Civil Engineering, IIEST, Shibpur, India
⁴ Dept. of Computer Science, IIT – BHU, Varanasi, India
⁵ IME, University of Chicago, Chicago, USA

Abstract— The world faces a grave water risk that affects all aspects of human life and ecology with implications for food security, energy production, industrial activity and human health. India is particularly affected as it has 16% of the world’s population but access to less than 4% of global freshwater resources. In this study, we use mobile (moving) sensors to spatially and temporally map river water quality based on in-situ data gathered in some of India’s major rivers. Data visualizations generated are intended to pinpoint sources of pollution, ensure regulatory compliance and examine health of the water body. We show that such cyber physical sensing techniques can be a powerful and more practical (or cost-effective) way to dynamically monitor, predict and regulate the quality of large bodies of water.

Keywords—Heatmaps, Water quality, Internet of Things(IoT)

I. INTRODUCTION

Water plays a vital role in sustaining the Earth’s ecosystem. United Nation’s Sustainable Development Goal 6 discusses challenges related to water scarcity and security [1]. Availability of water, both in terms of quantity and quality, helps determine its utility function to fulfill the growing demands of an increasing population. Enhanced industrial and agricultural activities also pose a threat to water resources in the form of pollution be it chemical, microbial or thermal [2, 3]. In order to have the defined utility value for any water source, one needs to take a close look at its quality parameters, which are important for its management. Without having reliable and consistent long-term water quality data, effective preservation and remediation won’t be possible for any water body [4]. Traditionally, field measurements for water quality evaluation have depended upon on-site sample collection, and transport to land-based or shipboard laboratories for evaluation, which is a costly as well as time and labour-intensive process [5]. Such methods have been limited in temporal and spatial scales to adequately address factors influencing the development of events such as harmful algal blooms, oxygen depletion and fish kills [6].

Recent advancements in technology in terms of mobile (moving) sensors can be leveraged for real-time online water quality monitoring to obtain a comprehensive picture [7]. While a stationary sensor-based approach would have ensured continuous monitoring, the costs of the large number of sensors required to monitor a given area do not justify the marginally greater insights that such data may present. Hence, a mobile sensor-based approach justifies lower costs and adequate data availability for a given area. The intent is to empower researchers and practitioners with such data and technology to capture the quality of river water using various parameters, in different seasons and at diverse locations across the river.

We discuss such an end-to-end Water-to-Cloud system that encompasses the gathering and curation of high resolution data on the Cloud in section II, explain our approach in detail in section III and present relevant initial results in section IV. Section V draws on important conclusions obtained so far and section VI briefly mentions our future plan of action.

II. RELATION WITH IOT AND CLOUD COMPUTING

Our novel approach lies in developing a scalable, low-cost, in-situ monitoring system using mobile sensing platforms carried around pre-selected routes to obtain high frequency temporally and spatially varying water quality data. In this study, we discuss results from the use of such systems across major rivers in India, namely Ganga, Yamuna and Godavari, based on non-stationary mobile sensors for mapping river water quality. Following curation of the geospatial and temporal water quality data, we intend to apply analytic and physics-based techniques to predict the spread of pollution, interpolate sparse data, and carry out inverse analysis to identify pollution sources.

III. METHODOLOGY

Figure 1 shows the general approach followed by our team for data collection, transmission and digital curation.

---

This study is sponsored by the Bill & Melinda Gates Foundation and the Tata Centre for Development at UChicago.
In section III A, tabulated information regarding the various parameters being monitored and their classification is shown. Section III B describes the sensors being used. Section III C provides insights on how the collected data is used and managed to achieve our main objective and section III D exhibits the process of data visualization and digital curation.

A. Parameters

A number of in-situ parameters being studied are listed in table 1. In order to have a complete understanding of all water quality parameters and to study correlations, ex-situ (lab based) method is employed in addition to in-situ testing.

Table 1: In-situ and ex-situ parameters monitored for this study

<table>
<thead>
<tr>
<th>In-situ measurement</th>
<th>Lab based measurement</th>
</tr>
</thead>
</table>

B. About the sensors in use

We have employed Hanna Instruments’ HI-9829 geotagged, time-stamped multi-parametric probes to monitor pH, turbidity, electrical conductivity, dissolved oxygen, temperature, salinity and nitrate ions. Turner Designs’ C3 instrument is being used to measure chlorophyll – a content, tryptophan and CDOM.

C. Working with data sets

Data collection
- Sensor based gathering of data
- Geo-tagged files in csv format

Data cleansing
- Noise removal by both manual approach and using the average moving technique.

Data Visualization
- geo-tagged csv files are used to generate the heatmaps to provide spatial and temporal information.

Analysis
- Correlations are determined between in-situ and ex-situ parameters to reduce reliance on labs.

D. Maps generation

In-house algorithms are run on data to create heatmap visualizations that are digitally curated on an open-source platform, https://thoreau.uchicago.edu.

IV. RESULTS AND DISCUSSION

We present results based on data gathered from April 2017 to July 2018 in various rivers across India, namely Ganga, Yamuna and Godavari. Our system’s unique ability to simultaneously examine various facets of river water pollution is demonstrated by the following results:

A. Spatial and temporal quantification of data sets

The visual representation of water quality parameters in the form of heatmaps helps in understanding the situation better as colour gradient variation with respect to the standard scale based on colour patterns is more intuitive for humans instead of large, technical tables of data, which restrict the analytical ability of a non-technical person.

In the figure 4, the variation in electrical conductivity in river Godavari around Rajahmundry on different days is illustrated to identify local pollution hotspots for each day. Alternatively, one could also compare variations for a parameter across different days via normalization.
C. Predicting Biological Oxygen Demand (BOD) and Chemical Oxygen Demand (COD)

Regression analysis and machine learning approaches are being examined for inferring BOD and COD values from parameters monitored in-situ. It has required the collection of large water sample sets that are measured both via laboratory measurements as well as in-situ measurements so that models can be trained on this data. Such an approach, if successful, would eliminate the need for cost-intensive and time consuming (3-5 days) tests in labs once the prediction models are validated.

D. Demonstrate effect of contaminant load on a water body

In order to conduct a one-week baseline study on the Yamuna, a 15 km route was identified at New Delhi. The flow of water from upstream is known to be regulated, and water is released once every few days, as can be seen in Figure 6 from left to right in terms of electrical conductivity. Fresh water was released a day before the start of the experiment and not again till the end of it. Increasing trend in EC values from day one to day four of experiments is indicative of the pollution load into the river. Foul smell increases progressively with incoming wastewater from drains daily. Such information could help in effective water management and can aid in creating 2D and 3D pollutant dispersion models.

E. Understanding concurrent effect of different parameters

Data visualizations help in demonstrating the association or disassociation of various parameters with each other, which helps in providing insights on how the river system functions in different dynamic situations. For example, during experiments at Narsapur in river Godavari, which is 10km from the sea, variation in Dissolved Oxygen (DO) levels in association with salinity and organic waste loading rate, which keeps on altering the river dynamics within the interval of few hours, highlighted the effect of wastewater drains when measured during morning or evening hours but failed to identify the same around noon due to high tides.

F. Defining water utility based on available data

Based on the concentration of pollutants at specific ghats in Varanasi or Rajahmundry, a layman could choose a safe site for the religious dip or swimming in these places. For example, Gauthami Ghat in Rajahmundry, a pilgrimage site associated with the holy dip, is often more polluted than nearby ghats due to high anthropogenic activities around it and should potentially be avoided by people with sensitive skin.

V. CONCLUSION

From the above results, it can be concluded that the Water-to-Cloud system can address key issues such as spatial distribution of pollutants, pinpointing contaminant sources, ensuring regulatory compliance for the concerned authorities, and benchmarking remediation measures; where traditional grab sampling method would have failed to capture these features due to its limitations. Moreover, the availability of raw data through an open source platform in near real time ensures transparency and accountability. Such data would contribute to effective policy discussions and ensure its use in decision making.

VI. WAY FORWARD

This study could be scaled up in order to understand the seasonal fluctuations in water quality for any river system. We plan to extend the representation of the water bodies in terms of deriving overall water quality indices for each river which could serve as a general method for easy comparison amongst different rivers and also facilitate quick actionable measures by relevant regulatory bodies. In order to predict a river’s health status in association with measuring both longitudinal and lateral mixing of the pollutants within the system and to develop an associated hydrological model for the prediction purpose in the future, collecting such high frequency spatial and temporal data is a key step.

Furthermore, we intend to validate our BOD and COD prediction models and create similar ones for total coliform and faecal coliform estimations over the next year.
REFERENCES


