Sustainable Environmental Monitoring via Energy and Information Efficient Multinode Placement

Sabtain Ahmad, *Graduate Student Member, IEEE*, Halit Uyanık, Tolga Ovatman, *Senior Member, IEEE*, Mehmet Tahir Sandıkkaya, *Senior Member, IEEE*, Vincenzo De Maio, *Member, IEEE*, Ivona Brandić, *Member, IEEE*, and Atakan Aral, *Senior Member, IEEE*

***Abstract*—The Internet of Things (IoT) is gaining traction for sensing and monitoring outdoor environments, such as water bodies, forests, or agricultural lands. Sustainable deployment of sensors for environmental sampling is a challenging task because of the spatial and temporal variation of the environmental attributes to be monitored, the lack of the infrastructure to power the sensors for uninterrupted monitoring, and the large continuous target environment despite the sparse and limited sampling locations. In this article, we present an environment monitoring framework that deploys a network of sensors and gateways connected through low-power, long-range networking to perform reliable data collection. The three objectives correspond to the optimization of information quality, communication capacity, and sustainability. Therefore, the proposed environment monitoring framework consists of three main components: 1) to maximize the information collected, we propose an optimal sensor placement method based on QR decomposition that deploys sensors at information- and communication-critical locations; 2) to facilitate the transfer of big streaming data and alleviate the network bottleneck caused by low bandwidth, we develop a gateway configuration method with the aim to reduce the deployment and communication costs; and 3) to allow sustainable environmental monitoring, an energy-aware optimization component is introduced. We validate our method by presenting a case study for monitoring the water quality of the Ergene River in Turkey. Detailed experiments subject to real-world data show that the proposed method is both accurate and efficient in monitoring a large environment and catching up with dynamic changes.**

***Index Terms*—Energy efficiency, environmental monitoring, gateway configuration, LoRaWAN, multiobjective optimization, QR decomposition, sensor placement, wireless sensor networks (WSNs).**

Manuscript received 13 March 2023; revised 19 June 2023; accepted 24 July 2023. Date of publication 8 August 2023; date of current version 7 December 2023. This work was supported in part by the CHIST-ERA-19-CES-005 Grant; in part by the SWAIN Project; Austrian Science Fund (FWF): I 5201-N; in part by the Rucon Project (Runtime Control in Multi Clouds); FWF: Y904-N31 START-Programm 2015; in part by the Triton project (Transprecise Edge Computing); FWF: P 36870-N; and in part by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grant 120N679. *(Corresponding author: Atakan Aral.)*

Sabtain Ahmad, Vincenzo De Maio, and Ivona Brandić are with the Institute of Information Systems Engineering, Vienna University of Technology, 1040 Vienna, Austria.

Halit Uyanık, Tolga Ovatman, and Mehmet Tahir Sandıkkaya are with the Department of Computer Engineering, Istanbul Technical University, 34469 Istanbul, Turkey.

Atakan Aral is with the Department of Computing Science, Umeå University, 901 87 Umeå, Sweden, and also with the Faculty of Computer Science, University of Vienna, 1090 Vienna, Austria (e-mail: atakan.aral@umu.se).

Digital Object Identifier 10.1109/JIOT.2023.3303124

I. Introduction

THE PROTECTION of aquatic and terrestrial environments is one of the key factors that play a crucial role in the sustainable growth of the whole world according to UN Sustainable Development Goals [1]. Health and hygiene are vital components of the sustainability of humanity, which comes from a clean and pollutant-free environment. Thus, its monitoring becomes essential to ensure the safety and well-being of society. Advances in machine learning (ML) and the Internet of Things (IoT) have laid the foundations for smart environment monitoring (SEM), enabling the means to sense, analyze, and act on the factors impacting the environment.

SEM provides the capability to tackle the challenges of high spatial and temporal scales. Recent developments in sensing technologies present an opportunity to efficiently monitor large geographical areas with high temporal frequency [2].

However, substantial challenges need to be overcome before the widespread deployment of sensors to collect data on large spatial scales. Sensor placement plays a vital role in SEM. Sensors must be placed in a minimal subset of locations yet allow the collection of the maximum amount of information on the region of interest [3]. Sensor placement in IoT applications is often handled by various placement strategies, such as random, uniform, polygon based, or grid based [4]. For a relatively moderately sized setting, optimal sensor locations are estimated by information theoretic tools, such as convex optimization [5], deep learning approaches [6], or Bayesian methods [7]. These approaches are unsuitable for an effective SEM since spatial information to describe the dynamic behavior of an environment or sensitivity of events of interest is not available in many locations [8].

Considering that only a limited number of sensors can be deployed because of various prohibitive factors (e.g., upfront and maintenance cost, ecological footprint), it is vital to deploy these sensor nodes at the most advantageous locations. High installation and energy costs coupled with the network limitations stemming from factors, such as obstructions (trees, buildings, etc.) and large distances result in inadequate spatial coverage and monitoring of the environment [9]. Furthermore, due to the nature of wireless communication, poor link quality caused by sensors that are too far apart or obstructed by trees or hills results in many retransmissions in an attempt to collect sufficient data. Packet retransmissions drain

battery power, a scarce resource, reducing the lifetime of IoT deployment [10].

The wireless sensor network (WSN) is a key enabling technology in an SEM application, which can be defined as a system that integrates a multitude of microcomponents, including IoT sensor nodes, gateways, and base stations [11]. Most work on sensor placement considers only one type of node, i.e., sensors or gateways [9]. However, in a resource-constrained environment where only a limited number of sensors and gateways can be deployed, the independent placement of sensors and gateways not only significantly affects the performance but also increases the cost of deployment. In addition, the impact of the number of gateways and the cluster of sensors assigned to each gateway on energy consumption is not thoroughly investigated, which could have considerable consequences on the overall deployment and performance [12], [13]. Finally, a significant part of the previous work considers sensor placement only for a small region of interest, in which the impact of the distance between sensor nodes is not prominent and, thus, excluded in placement models [14].

The goal of this work is to propose an SEM method based on multinode (sensor and gateway) placement, incorporating heterogeneous constraints. The below objectives are identified.

- OBJ1 Information quality, i.e., the extent measured data represents the target phenomenon. The data quality can be maximized by choosing the sampling locations strategically based on the spatial distribution of the measured parameters.
- OBJ2 Communication capacity, i.e., the amount of data that can be reliably transferred from the sensors to the processing location. The sensors need to communicate with the gateways to transmit the collected data. In large outdoor environments, this connection between sensors and gateways is generally affected by long distances and obstacles.
- OBJ3 Sustainability, i.e., the ecological footprint, energy efficiency, and economic viability of the measurement infrastructure. A sensor undertakes different actions, such as transmitting/receiving data, processing, and remaining in an idle state with different energy consumption characteristics, whereas the energy consumption of a gateway depends on its signal range.

We aim to determine how can an IoT deployment achieve these three objectives, and make the following contributions.

1) We propose a unified environment monitoring framework consisting of four components that allow an estimate of a certain spatiotemporal environmental state. Unlike most previous work, where only a single type of node is considered for tackling the sensor placement problem, in this work, we study the interdependence of sensor and gateway placements. We use LoRaWAN-based communication due to its low-cost, low-power characteristics, and long-distance capability. Moreover, generated placements are on a continuous space instead of conventional but rigid structures, such as polygons or grids.
2) We model the energy consumption of the proposed candidate deployments consisting of different numbers and configurations (e.g., range) of sensors/gateways and aim at building a robust and energy-efficient measurement infrastructure. To the best of our knowledge, no such evaluations exist for any other SEM method based on multitype sensor nodes.
3) We validate the proposed algorithms within the context of the sustainable watershed management through IoT-driven AI (SWAIN) project using real-world water quality data collected from 75 locations along the Ergene River, Turkey.

TABLE I
NOTATION USED IN THIS WORK

Variable	Definition
m	number of samples
$\mathbf{x_i} \in \mathbb{R}^{\mathbf{n}}$	n-dimensional signal
$\boldsymbol{X} \in \mathbb{R}^{m \times n}$	samples matrix
$\mathbf{e}_\gamma \in \mathbb{R}^{\mathbf{n}}$	standard basis vector
$\boldsymbol{C} \in \mathbb{R}^{r \times n}$	canonical basis matrix
$\mathbf{S}$	an index set of sensor locations
τ	cost vector
NCC	network communication cost
GTC	gateway cost; distance to the nearest gateway
SS	site score
$E(s)$	energy consumed by sensor s
$E_{passive}(s)$, $E_{active}(s)$, $E_{comm}(s)$	respectively, energy for passive, active and communication state of s
$P_{idle}(s)$	Idle power of sensor
t	time instant for power integration
θ	distance threshold between sensor and closest gateway

For the rest of this article, we first provide the background and related work on the techniques used in this work, as well as our use case scenario in Section II. Then, in Section III, we explain our solution, the GENS framework [(G)ateway placement, (E)nergy modeling, (N)etwork modeling, (S)ensor placement], in detail. We present an analysis and experimental evaluation of GENS using data from a real-world water quality monitoring system in Sections IV and V, respectively, and conclude the article in Section VI. Table I presents the common notation used throughout this article.

II. BACKGROUND AND RELATED WORK

A. Use Case Scenario: Water Quality Monitoring

Water quality monitoring networks (WQMNs) are projected to cover the life-cycle of freshwater usage from spring to wastewater [15]. This broad spectrum implies many parameters, including weather conditions, ecology, industrialization, and urbanization [16], [17]. Proposed WQMN: 1) performs continuous water quality monitoring to collect quantitative chemical properties of the surface water in a river basin and 2) could sense the fluctuations in the chemical composition of the water for detecting pollution rate or temporal discharges [18]. Such WQMN with accurate and timely measurements requires to have sophisticated sensors at least on each tributary. This is a high-cost solution as a large river basin can have thousands of tributaries. Also, access to most tributaries can be drudging due to several climatic and

topographic obstructions at each altitude. Moreover, managing, communicating with, and providing power to high-end devices while leaving them unattended in the wild is challenging. Therefore, a dependable yet low-cost network must optimize not only the sampling locations but also the reliability of the collected data. An optimized network must be able to autonomously gather relevant data from decisive sampling points, efficiently and reliably transmit timely data to a management system, and extend the lifetime of the whole network for continuous interpretation of the measurements.

B. Low Power Wide Area Network Communication

In order to establish the link between IoT sensors and the gateway units in a wide-range rural communication scenario while also minimizing the energy consumption rates at the cost of possibly having lower data rates (DRs), different low power wide area network (LPWAN) protocols and technologies, such as Sigfox[1], random phase multiple access (RPMA)[2], NB-IOT[3] and LoRaWAN, can be used [19], [20], [21], [22], [23]. Across the different technologies, although the chosen LoRaWAN protocol does not outperform on every aspect of LPWAN communication, it provides a balanced efficiency between latency, coverage area, and payload length while also being cheaper in terms of gateway deployment [20].

Sigfox runs in unlicensed bands ranging from 862 to 928 MHz. Its uses ultra narrowband (UNB) modulation to support up to 140 uplink message per day with a maximum of 12-byte payload size and a maximum data-rate of 100 bps. It claims to have 40-km range outdoors [19], [20]. RPMA aims to scale up to billions of IoT devices while also maximizing the wireless device life to last as long as possible. According to its official documentation, with only around 20 towers, it can provide coverage up to 2000 miles while minimizing the interference from obstructions, such as walls and agriculture. With high link budget that operates at 2.4-GHz spectrum, 21 and 30 transmission power (TX) for uplink and downlink communication, it claims to have better coverage than the existing LPWAN protocols [20], [21], [23]. NB-IoT uses licensed LTE frequency bands which introduce additional cost to the users for spectrum usage unlike Sigfox and LoRaWAN. It provides up to 1600-bytes payload size and can support up to 10-km range in rural areas with 200-KHz bandwidth. Unlike Sigfox and LoRaWAN, it is not as interference resistant. It does not use adaptive data-rate and since it uses additional communication steps to ascertain certain QoS conditions, it ends up using more energy than protocols such as LoRaWAN [20], [21].

LoRaWAN is a low-power wide-area network communication protocol that is defined on top of the LoRa modulation technique [24]. LoRa is patented by Semtech,[4] and the specifications of LoRaWAN is defined by the LoRa alliance.[5] The protocol uses unlicensed Industrial Scientific Medical (ISM) bands while operating under the duty-cycle and power regulations defined by local institutions. In the context of this study, a gateway unit is used as a sink and sensors as data generators in order to establish LoRaWAN communication.

LoRaWAN is useful in rural areas for enabling data collection for many different purposes, such as health monitoring, fire alarms, and agricultural support [25], [26], [27], [28]. An important issue within these environments is the gateway placement and corresponding cost problems, which have been studied widely in the literature. Some solutions formulate the issue of gateway and sensor placement as an optimization problem and try to find optimal locations while considering important LoRa parameters, such as spreading factor (SF) and power consumption [29]. As [30] mentions, it is possible to utilize many different optimization, clustering, and genetic algorithms to figure out different gateway placement approaches with respect to different metrics. When the communication needs to respect QoS requirements, gateway placement should be done while considering different metrics [31]. Other studies focus on reducing collisions within the network, which is a massive issue in LoRa-based networks, especially those that cover long ranges [32]. By reducing both collisions and the number of gateways, it is possible to reduce the overall cost of network deployment. Another issue that is not within the coverage of this study but is important is the scalability of the gateway placement, which needs to support future possible network congestion when the number of sensors increases [33].

In the proposed architecture, the sensors are distributed in a vast rural area across a large river basin. This environmental setup demands long battery life as well as long-span communication with minimum cost on gateway placements for sensors. The LoRaWAN protocol provides these required aspects by design. Another advantage of using LoRaWAN in our setup is the ability to utilize a limited number of frequency channels in rural areas with fewer collisions.

C. QR Decomposition

Within the field of SEM, the sensor placement problem has been studied extensively in indoor environments, where spatial distance is not of crucial importance [11], [34]. Therefore, the studies tend to focus on a single objective: maximizing the information collected. For vast outdoor environments, the goals of sensor placement are different and, thus, different sensor placement strategies, considering multiple constraints, are required. Determining optimal number of sensor to be placed is intractable via brute-force search among the combinatorial possibilities. There are $\binom{n}{k} = (n!/k!(n-k)!)$ possible choices of k sensor nodes out of an n-dimensional space.

Monitoring a vast outdoor environment only with a handful of sensors requires considering the cost and data loss. When the area to be monitored is relatively large, the cost of prevalent sensor placement would be very high and may result in transmission delay and slow response. The data corresponding to the whole region can be reconstructed with a small amount of information collected through sparse sensors if the spatial correlation of the monitoring information is high. When

[1]https://www.sigfox.com/
[2]https://www.ingenu.com/technology/rpma/
[3]https://www.3gpp.org/
[4]https://www.semtech.com/lora
[5]https://lora-alliance.org/about-lorawan/

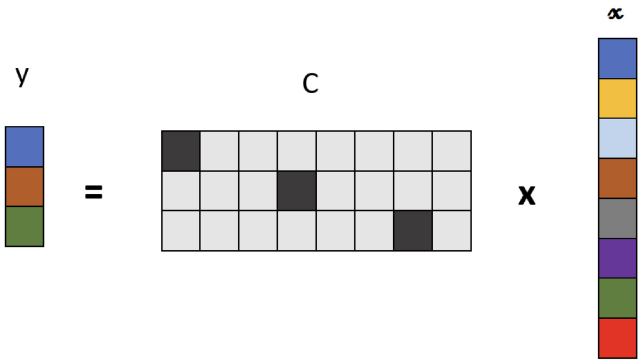

Fig. 1. Sparse sampling example: locations (1st, 4th, and 7th) are extracted according to $\boldsymbol{C}$ in which black square represents 1, and the rest of the matrix contains 0.

sensors are sparse, it is essential to study the deployment methods to enhance the accuracy of the reconstructed data. Therefore, it is desirable to determine optimal locations to achieve the highest reconstruction accuracy.

We model the sensor placement as a reconstruction problem, that is, reconstructing a high-dimensional signal from a limited amount of in site measurements. The sampling locations are optimized to extract sufficient features for reconstruction and prediction from the input space. Let $\mathbf{x_i} \in \mathbb{R}^{\mathbf{n}}$ denote an n-dimensional spatiotemporal signal and $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ denote a matrix that contains the samples $\mathbf{x_i}$. Then the measurement matrix $\boldsymbol{C} \in \mathbb{R}^{r\times n}$ can be used to select the sampling locations

$$\boldsymbol{C} = \left[\mathbf{e}_{\gamma\mathbf{1}}, \mathbf{e}_{\gamma\mathbf{2}}, \ldots, \mathbf{e}_{\gamma\mathbf{r}}\right]. \tag{1}$$

Here, $\mathbf{e}_\gamma \in \mathbb{R}^{\mathbf{n}}$ are the standard basis vectors that are "1" at γ_{th} element and $r \leq m$. An example of sparse sampling is shown in Fig. 1. The measurements at sampling locations $\mathbf{y}$ are chosen to guarantee the best feasible reconstruction of $\hat{\boldsymbol{X}}$. The sampling locations in the signal correspond to in-situ measurements in $\boldsymbol{X}$. A data-driven approach such as singular value decomposition (SVD) can be used to find the optimal sensor locations and the suitable canonical matrix $\boldsymbol{C}$ [35]. Given a suitable canonical basis matrix $\boldsymbol{C}$ the data matrix can be compressed as shown in

$$\boldsymbol{Y} = \boldsymbol{CX}. \tag{2}$$

Only keeping r basis from SVD results in $\boldsymbol{T_r} = \boldsymbol{XV_r^T}$. Therefore, the spatiotemporal signal $\hat{\boldsymbol{X}}$ can be represented using sparse samples

$$\hat{\boldsymbol{X}} = \boldsymbol{T_r}(\boldsymbol{CT_r})^{-1}\boldsymbol{Y}. \tag{3}$$

As shown above, the optimization of sensor locations also requires the optimization of the measurement matrix. Therefore, the sensor locations are optimized by maximizing the singular value spectrum of the principal components

$$\hat{\mathbf{S}} = \arg\max_{S} |\det \boldsymbol{C_s T_r}|. \tag{4}$$

The optimization of (4) can be performed by greedy heuristic approaches such as empirical interpolation methods (EIMs) [36]. In this work, we employ the pivoted QR decomposition-based approach for the optimization of (4). The optimal sensor locations correspond to the top r pivots of the QR decomposition

$$\left|\det \mathbb{A}\mathbb{C}^{\mathbb{T}}\right| = |\det\mathbb{Q}||\det\mathbb{R}| = \prod_i |r_{ii}|. \tag{5}$$

D. Multiobjective and Multinode Sensor Placement

Sensor placement in WSN presents a hard optimization problem, which is further complicated by the need to take gateway locations into consideration and the requirement for energy minimization and network lifetime maximization. A WSN consists of a large number of sensors and gateways (sink nodes) to monitor a target area. The locations of sensors and gateways should be determined to maintain low cost, network connectivity, and coverage. The multiobjective sensor placement optimization methods can be broadly classified into two categories: 1) the use of weight factors to realize different combinations of fitness functions and to essentially transform them into a single objective and 2) the use of a multiobjective optimization algorithm to find a tradeoff between conflicting objectives by generating Pareto front and nondominated solutions. A multiobjective strategy based on information entropy is developed in [37] to detect damages in bridges. A multiobjective combinatorial optimization based on a reduced order model is proposed in [38] to maximize the monitoring performance while minimizing the deployment cost. A genetic algorithm-based sensor placement optimization algorithm is presented in [39], which results in significant improvement in structural damage detection accuracy compared with traditional approaches. Energy-aware sensor placement is critical for network lifetime as these sensors are inherently resource-constrained [40]. A stochastic EfI method [41] is presented for determining sensor locations based on optimal energy usage.

Another factor that influences the reliability and efficiency of the deployed network is the gateway locations [42]. Gateway locations can affect the robustness of communication and network lifetime as sensors need to be connected with the appropriate gateways to ensure robust and efficient data collection. Furthermore, the distance between the sensors and the gateways is one of the most important factors affecting energy consumption. To address the problem of positioning gateways, researchers have formulated this as an optimization problem and attempted to solve it using metaheuristic methods [43], ant colony optimization [44], and particle swarm optimization algorithms [45]. Several other techniques have also been reported in the literature, for instance, the p-median model to determine gateway locations [46] and energy-aware clustering architecture for maximizing network lifetime [47].

The research reviewed in the literature suggests that the multiobjective and multinode placement problems have been solved with relatively small to medium-sized networks to monitor small to medium target areas. Moreover, the independent optimization of sensor and gateway placements can result in suboptimal performance [48]. Therefore, a notable gap exists in this field, where the optimization of both sensor and gateway placements needs to be addressed, considering the limited network availability, variable placement costs, and objectives related to information quality and energy efficiency.

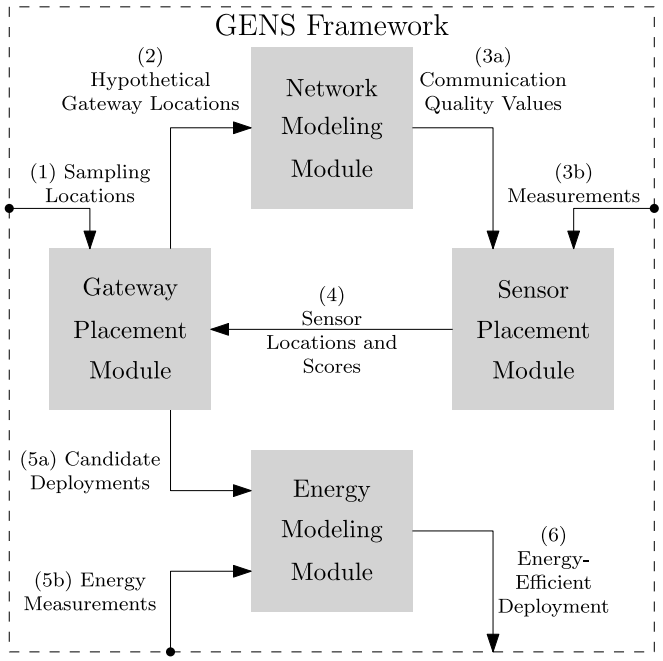

Fig. 2. Data flows among the four modules of the GENS framework. The external inputs are highlighted with dots.

In this work, the main goal is to monitor water quality; therefore, objectives, such as information quality, have to be prioritized over others (i.e., network latency and energy efficiency). Aggregation approaches, such as [49], are widely used to prioritize objectives. However, it is difficult to determine the correct weights to achieve the best solution. To this end, instead of finding a Pareto-optimal solution for all identified objectives, we perform a multistage optimization, similar to [50]. In this class of approaches, first, an optimal solution for a specific objective is found. Then, the solution is updated in subsequent iterations to improve it also for other objectives.

III. GENS Framework

The four modules that constitute the GENS framework, their order of execution, and input/output relationships among them are demonstrated in the data-flow diagram in Fig. 2. These modules contribute to achieving OBJ1–3 introduced in Section I. More specifically, the process starts with the gateway placement module, which takes the offline sampling locations (1) as its input and generates a hypothetical gateway deployment (2). In this deployment, it is assumed that each offline sampling location is equipped with a sensor. Although far from optimal in terms of OBJ1 and OBJ3, the hypothetical deployment provides the means of estimating the optimal communication quality (OBJ2) for a large number of locations covering the target environment. The estimates (3a) are based on the network simulation within the network modeling module. Moreover, the use of initial sampling locations enables us to incorporate use cases with a monitoring infrastructure already available, which can be improved upon through the GENS framework.

In the next step, the sensor placement module identifies the optimal locations for collecting the most representative data (toward OBJ1) based on offline measurements (3b) and for transferring them effectively (toward OBJ2) based on communication quality (3a). The output is the optimal sensor locations ranked by their scores (4), which the gateway placement module utilizes to determine the actual gateway locations. This module ensures that streaming data from these sensor locations can be transmitted to where they are processed (toward OBJ2). At this point, it is possible to avoid an excessive number of sensors and gateways (toward OBJ3); however, the actual impact on energy consumption is not known. For instance, a higher number of gateways could result in lower consumption due to the reduced distance of LoRaWAN data transmission. Therefore, several candidate solutions with varying gateway and sensor counts or gateway coverage areas are generated.

The purpose of the energy modeling module is to evaluate these solutions (5a) using energy consumption models that rely on real measurements (5b) and determine the final energy-efficient deployment (6) toward OBJ3. The GENS framework facilitates the cost-effective monitoring of remote outdoor environments, such as forests or rivers, that are characterized by a lack of network availability and a dependable energy source. This is achieved through the deployment of sensors and gateways at critical locations, which optimizes network utilization and energy consumption, without compromising the information quality. It is worth noting that the proposed module implementations can be easily replaced by alternative algorithms without changing the high-level GENS framework.

A. Gateway Placement

The gateway placement module utilizes sensor scores (Section III-C) and network simulation (Section III-B) to determine the placement with the highest communication quality that also covers the sensors with the highest scores. To that end, we use a clustering algorithm to associate the sensors. The first step is calculating the signal range for each sensor represented by a sphere. This radius is an input parameter that ranges between 500 and 10 000 m. Distance ranges are chosen in a wide spectrum to observe the negative effects of distance on communication performance. Then, these spheres and their intersections with the earth's surface are calculated. These intersections are considered potential gateway locations.

An example of determining the intersection and potential gateway locations can be seen in Fig. 3. When a gateway is placed in such an area, all sensors covering the intersection will be able to communicate with the corresponding gateway. However, the number and area of intersections vary with signal range assumptions. In order to decide on the intersections to be used as gateway locations, they are sorted in descending order by the number of sensors they cover. Then, they are added to the selection iteratively until all sensor set is covered. The final gateway location is determined as the geographical midpoint of the intersection points, whereas the gateway altitude is the average height of the sensors within their respective cluster.

Aside from the signal range and sensor score threshold, two more parameters are used during the clustering process. The first is a threshold that removes the clusters with fewer sensors than the given value. The value 0 indicates that no

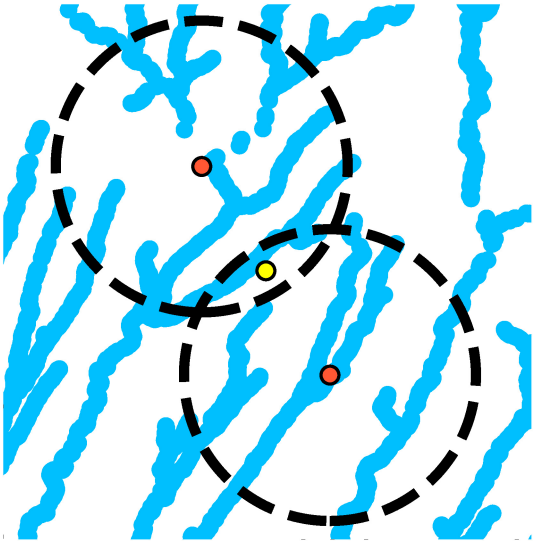

Fig. 3. Clustering example. The intersection of the signal ranges (dashed circles) for the river sensors (red dotes) is used to determine potential gateway (yellow dot) regions.

TABLE II
CLUSTERING & NETWORK SIMULATION PARAMETERS

Parameter	Values
Frequency sub-bands	868-868.6, 868.7-869.2, 869.4-869.65
Path-loss exponent	2.2
Reference distance	100m
Reference loss value	78dB
Data-rate	DR0
Spreading factor	SF12
Bandwidth	125kHz
Number of sensors	10-75
Gateway cluster range	500m-10000m
Sensor duty-cycle duration	135s
Simulation duration	86400s

clusters are removed. This parameter can be used to reduce the cost of deployment. The second parameter is a binary value that determines whether the sensors within multiple gateway ranges should be removed, reducing the possible signal conflicts during the network simulation process.

B. Network Modeling

A simulation environment is created using ns-3 network simulator [51] to evaluate the communication quality with the calculated gateway locations. ns-3 does not ship a built-in LoRaWAN support; therefore, a third-party LoRaWAN module is used [52]. This module uses class-A device type defined under LoRaWAN specifications and its corresponding communication flow, which has the lowest power consumption amongst device classes A, B, and C [53]. We expand the default state machine model used by this module to reason about the hardware properties we use in our simulations. The expanded state machine manages the state changes of the sensors to schedule the packet transmissions to the gateways.

Important parameters used in the clustering and network simulation process are given in Table II. For the input parameters path-loss exponent, reference distance, and reference loss values fed into the network simulation, it is important to consider that different obstacles, specifically when the sensor antenna is placed lower than the tree height level, will affect the network performance. Therefore, these values for calculating the signal power are chosen as 2.2, 100 m, and 78 dB, respectively, in accordance with a previous study that focuses on different vegetation densities [54]. These values correspond to low-density tree vegetation in accordance with the selected experimental use case (see Section IV).

Additionally, three sub-bands are used by the LoRaWAN module; 868–868.6 and 868.7–869.2 for uplink and downlink requests with the duty cycle of 1% and 0.1%, respectively, and 869.4–869.65 with a duty-cycle of 10% for downlink requests within the second receive window of class-A LoRaWAN device communication. This enables the gateway to send more ACK packets compared to using only 1% duty-cycle frequencies. When it is not possible to send a packet within the first receive window due to the duty-cycle limitation of 868.1 frequency, gateways can send ACK packets within the second receive window using the higher 868.5 frequency. Therefore, the base 868.1 frequency used for uplink messages and its corresponding duty-cycle is split into equal time slots for each sensor. Another frequency with a higher duty-cycle for downlink can be utilized by the gateway to send back ACK messages. It should be noted that when the gateway is in its downlink transmit mode, it cannot receive the incoming packets. The adaptive data-rate approach is not utilized, and due to the large distance covered by the clusters, the DR is set to 0, and the SF is set to 12 by the LoRaWAN module.

Our implementation also includes geographical details, detailed hardware energy consumption rates, and simulation parameters. Due to space limitations, parameters set to values from the literature are not discussed here. Interested readers may refer to the source code repository,[6] which explains the simulation and clustering setup in further detail.

C. Sensor Placement

The sensor placement model consists of a framework that scales to arbitrarily large problems, leveraging modern techniques in ML. Reducing the number of sensors through sensor placement can reduce the deployment cost and enable faster estimation of the environment. For the cases where sampling resources are limited, the deployed sensors may not be able to monitor the environment effectively as they fail to capture all relevant features. To address that, we develop a framework to reconstruct and predict the entire spatiotemporal signal from limited samples. Data-driven approaches such as QR decomposition can be used to optimize sensor placement [55]. We propose an MO sensor placement method for signal reconstruction based on column-pivoted QR decomposition (MSPQR). Compared to model-based optimization methods, data-driven approaches have lower complexity, are easier to implement, and scale better to large areas.

Data from physical systems typically possess an extremely low-rank spatiotemporal correlation structure that can be exploited to reduce the number of sensors for global inference drastically. Consider a matrix $\boldsymbol{X}$ of m samples of data

[6]https://github.com/HalitU/SWAINlorawangateways

$\mathbf{x_i} \in \mathbb{R}^{\mathbf{n}}$ in which the columns correspond to spatial locations

$$X = \begin{pmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \vdots \\ \mathbf{x_m} \end{pmatrix}. \tag{6}$$

Furthermore, assume ψ to be a matrix derived from X (e.g., linear combinations of rows or right singular vectors of X). For an index set $\mathbf{S}$, a matrix X_s formed by collecting the columns of X with indices in $\mathbf{S}$. The set of indices $\hat{\mathbf{S}}$ that maximizes the determinant of ψ_S or product of singular values of ψ_S provides optimal interpolation points [56].

Even though finding such a set of indices $\hat{\mathbf{S}}$ is nonconvex and NP-hard, there are some reasonable approximate algorithms [57], [58]. However, these approaches do not scale well with data dimensionality. The algorithm for placing sensors at optimal and cost-effective locations for environmental monitoring by maximizing the product of the singular values of ψ_S is based on pivoted QR decomposition. The pivoted QR algorithm is more scalable and substantially more efficient than near-optimal and polynomial-time solutions. As explained previously, a high-dimensional signal can be reconstructed from sparse samples through sparse sampling techniques

$$\hat{\mathbf{T}} = \arg\min_{T} ||X - X_S T||_F. \tag{7}$$

where $||.||_F$ is the Frobenius error norm, and X_S is the matrix consisting of columns of X with indices in $\mathbf{S}$. The solution for (7) is the least-squares solution (Moore–Penrose pseudoinverse). Therefore, the reconstruction error is given by

$$e(\mathbf{s}) = \frac{||X - X_s X_s^{\dagger} X||_F}{||X||_F} \tag{8}$$

where $X_s^{\dagger}$ is the Moore–Penrose pseudo-inverse of X_s.

Including cost constraints in sensor placement problems is particularly important for monitoring large environments where some sensor locations cost more than others but may be more informative. Given the cost vector τ associated with each sample location and the budget b, the objective function, which also includes the cost constraint, is defined as

$$\hat{\mathbf{S}} = \arg\min_{S} e(S) \quad \text{s.t.} \sum_{s \in S} \eta_s \leq b. \tag{9}$$

For a given b, there exists a λ such that

$$\hat{\mathbf{S}} = \arg\min_{S} e(S) + \lambda \sum_{s \in S} \eta_s. \tag{10}$$

The sensor placement is modeled as a cost-constrained problem in a relaxed form. A parameter for tracing out the cost-error curve is introduced, representing the balance between reconstruction quality and cost. The algorithm for finding optimal sensor placements is based on a modification of pivoted QR decomposition [55]. The intuition for using column pivoted QR decomposition for sensor placement is that, for a given k, the first k pivots should be a suitable choice of sensor locations. Pivoted QR decomposition approximates the optimal solution by iteratively updating $\mathbf{S^{k+1}}$ so that $|\det A^{k+1}|$ is maximum. This process is summarized in Algorithm 1. Let

Algorithm 1 Sensor Placement: Column Pivoted QR Decomposition

Input: data matrix X, Q, R, number of sensors k, communication cost vector τ, cost-error balance γ
Output: set of k optimal locations **s**, site scores

```
S ← 1:n
for j = 1, ..., k do
    for i = 1, ..., n do
        v_i ← ||R_{j:m,i}||_2 − γ τ si
    end for
    l ← index of the maximum of v_{j:n}        ▷ select pivot
    v ← R_{j:m,j−1+l}
    swap(R_{j:m,j}, R_{j:m,j−1+l})
    swap(s_j, s_{j−1+l})
    σ ← ||v||_2
    u ← (v + sign(v_1) σ e^1)/√(2σ(σ + |v_1|))
                        ▷ normalized Householder reflector
    R_{j:m,j:n} ← R_{j:m,j:n} − 2uu^T R_{j:m,j:n}
    Q_{:,j:n} ← Q_{:,j:n} − 2Q_{:,j:n} uu^T
    SS_j = RS * (1 − τ)
end for
```

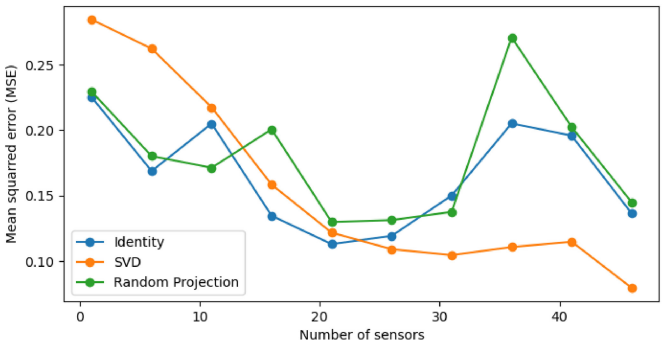

Fig. 4. Mean squared signal reconstruction error.

X be the data matrix containing m samples of an n-dimensional signal, R be the projections (i.e., random) of X, and Q as an identity matrix.

1) Model Training: The sensor placement model based on pivoted QR decomposition involves a few hyperparameters, such as the type and number of the basis used for matrix reconstruction [55]. These hyperparameter values are chosen by experimenting with the different combinations of values using the grid search approach. However, apart from these hyperparameters, some other parameters, such as sensor range and type of cost used to specify the cost constraints, are not directly linked with the model but influence its performance. The effect of these parameters is studied and explained in Section V-B. To assess the reconstruction capability of the model, the data set is divided into a training set and a test set. The output consists of a ranked list of optimal sensor locations with corresponding scores. The number of optimal locations influences the reconstruction error.

Fig. 4 demonstrates the relationship between the number of sensors and reconstruction error. To extract the tailored basis for our data set, we adopt the methodology used in [55] to

select the type and number of basis before applying the pivoted QR decomposition. Of the three preprocessing methods (SVD modes, randomized modes, and the raw data), SVD yields the best results, as is evident from Fig. 4. The model's ability to reconstruct the signal enhances as we increase the number of sensors. However, due to the cost and energy constraints, we limit the number of sensors to 30, for which the reconstruction error is sufficiently low. This corresponds to 60% reduction in the deployed sensors and significant energy and cost savings.

D. Energy Optimization

Finally, we apply energy optimizations to improve the energy efficiency of placements found in previous steps.

1) Energy Model: First, we develop an energy model based on energy consumption data collected through network simulation. We assume all communication is performed through LoRaWAN protocol, as it is common in the target scenario. For each sensor s in a placement configuration P, energy consumption $E(s)$ is equal to the sum of its *passive* energy $E_{\texttt{passive}}$, i.e., energy consumed for basic sensor operations (idle and low power operational modes) plus the *active* energy (i.e., for processing and active sleep), $E_{\texttt{active}}$, and energy for communication, $E_{\texttt{comm}}$. Being energy the integral of power over time, we define $E_{\texttt{passive}}(s)$ and $E_{\texttt{active}}(s)$, respectively, in (11) and (12), where t is the current time instant

$$E_{\texttt{passive}}(s) = \int_{\texttt{start}}^{\texttt{end}} P_{\texttt{idle}}(s,t) + P_{\texttt{sleep}}(s,t)\ dt \quad (11)$$

where $P_{\texttt{idle}}$ is the power the sensors in idle mode consumes

$$E_{\texttt{active}}(s) = \int_{\texttt{start}}^{\texttt{end}} P_{\texttt{active}}(s,t) + P_{\texttt{trans}}(s) + P_{\texttt{comm}}(s)\ dt \quad (12)$$

where $P_{\texttt{trans}}$ is the power consumed during the power-level transitions (i.e., from sleep to active mode), $P_{\texttt{active}}$ is the power consumed while the sensor is in active mode, and $P_{\texttt{comm}}$ is the power consumed during communication. Since $P_{\texttt{idle}}$, $P_{\texttt{sleep}}$, $P_{\texttt{active}}$ and $P_{\texttt{trans}}$ are constant over time, we focus on communication energy, $E_{\texttt{comm}}(s)$. Since we are interested in the energy consumption of a specific placement, we include in the energy of communication per sensor $e_{\texttt{comm}}(s, g_s)$ also the energy consumption for data transfer of its closest gateway. $E_{\texttt{comm}}(s)$ is the sum of each data transfer between the sensor and its closest gateway, g_s. We define the energy of a data transfer between a sensor s and gateway g_s $e_{\texttt{comm}}(s, g_s)$, as in

$$e_{\texttt{comm}}(s, g_s) = \int_{\texttt{start}}^{\texttt{end}} P^{s}_{\texttt{comm}}(s, g_s, t) + P^{g}_{\texttt{comm}}(g_s, s, \theta(g_s), t)dt \quad (13)$$

where θ is the distance threshold within which the sensor communicates and $P^{s}_{\texttt{comm}}(s, g_s, t)$ and $P^{g}_{\texttt{comm}}(g_s, s, \theta_{g_s}, t)$ are, respectively, the power consumption of communication for sensors and gateways. In $P^{g}_{\texttt{comm}}(g_s, s, \theta_{g_s}, t)$, we consider θ_{g_s} since the TX of gateway depends on the geographical distance from the sensor. Also, `start` indicates the time instant where the transfer starts, and `end` is calculated according to the amount of data and the available bandwidth, i.e., `end` = `start` + $[data/bw(s)]$. $E_{\texttt{comm}}(s)$ is then equal to the sum of the energy of each transfer between s and g_s, $e_{\texttt{comm}(s,g_s)}$. Energy consumption for placement P is $E(P)$ in

$$E(P) = \sum_{s\in P(S)} E(s) = \sum_{s\in P(S)} E_{\texttt{passive}}(s) + E_{\texttt{active}}(s) + E_{\texttt{comm}}(s) \quad (14)$$

where $P(S)$ is the set of sensors included in placement P. For our placement optimization, we focus on minimizing $E(P)$.

2) Energy-Aware Placement Optimization: The method used to reduce the energy consumption of gateway placement is described in this section. We assume that the placement solution contains a list of gateway and sensors. While sensors' and gateways' locations are fixed by sensors' and gateways' placement modules, gateways' LoRaWAN modules can be in different admissible configurations, which affect $E_{\texttt{comm}}$. The goal of this module is to find a configuration in the proposed gateway placement that reduces $E_{\texttt{comm}}$.

Algorithm 2 describes the generation of the initial population. We start to explore alternatives based on a starting solution computed by previous models, which we refer to as *pivot* solution. Starting from pivot solution p, nP different solutions are generated by randomly selecting an alternative threshold for each gateway in the pivot solution (lines 4–9).

The energy optimization is described by Algorithm 3. First, we calculate $E_{\texttt{comm}}$ value on one day-long execution, based on (14) (Algorithm 3, line 1). Then, we select the solution with the lowest $E_{\texttt{comm}}$, which will be the *pivot* solution for generating the initial population (lines 3–9). Once population P has been generated, we explore possible solutions by applying genetic algorithms metaheuristic in lines 11–19 [59]. First, we calculate the *fitness* of each solution in population, i.e., $E_{\texttt{comm}}$. Afterward, we apply a uniform crossover with binary tournament selection to allow the combination of the solutions with lower fitness. Then, we apply mutation to allow further exploration of solution space. At the end of this phase, we select the best *pSize* solutions to continue with further iterations until the termination condition is reached. In our case, we consider a termination condition either reaching the maximum number of iterations, *nIter*, or if the minimum $E_{\texttt{comm}}$ value does not change in the last ten iterations.

Algorithm 2 Generation of Population

Input: *pivot*, *pSize*
Output: P

```
1:  P ← ∅
2:  for i ∈ [0, pSize] do
3:      s ← ∅
4:      for all g ∈ pivot do
5:          thresList ← possibleThres(g)
6:          k ← rand(0, length(confList))
7:          g' ← confList[k]
8:          s ← s ∪ {g'}
9:      end for P ← P ∪ {s}
10: end for
11: return P
```

Algorithm 3 Energy Optimization [59]

Input: P, $pSize$, $nIter$, $condition()$
Output: $P'[0]$

$minEnergy \leftarrow computeEnergy(P[0])$
$pivot \leftarrow P[0]$
for all $p \in P$ **do**
  $tmpEnergy \leftarrow computeEnergy(p)$
  if $tmpEnergy \leq minEnergy$ **then** ▷ Find pivot
    $minEnergy \leftarrow tmpEnergy$
    $pivot \leftarrow p$
  end if
end for
$iter \leftarrow 0$
$P' \leftarrow generateConfigurations(p, nP)$
while $iter < nIter \wedge condition()$ **do**
  $fitness(P')$
  $crossover(P')$
  $mutation(P')$
  $P' \leftarrow selection(P', nP)$
  $i \leftarrow i + 1$
end while
return $P'[0]$

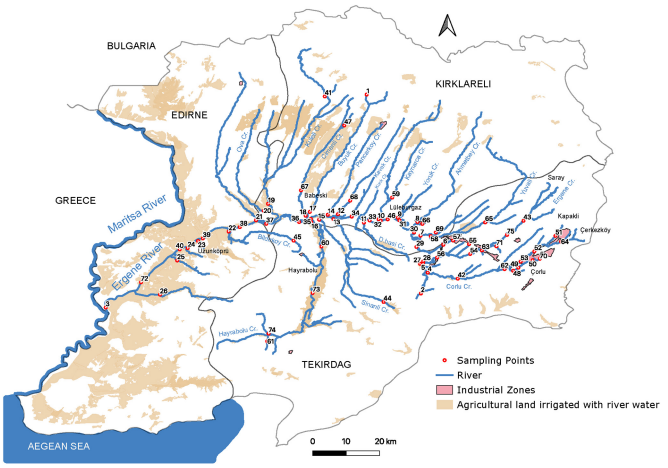

Fig. 5. Geographical overview of the Ergene watershed use case located in Northwestern Turkey (the image is courtesy of TUBITAK project 115Y064).

IV. Use Case Analysis

The presented study is applied for water quality monitoring of the Ergene River located in the Thrace region of Turkey, which is shown in Fig. 5. The environment intended to monitor for micropollutants in this case study is stretched to 420 km. The total catchment area is 14 439 km^2. The area is heavily industrialized, and about 40% of the catchment is active agricultural land. It is a very complex site for environmental monitoring, as the anthropogenic disturbances include industrial discharges and discharges from the water treatment plants. All these episodic changes dictate the chemical and physical parameters at the river and, thus, increase its complexity. A large amount of activity and its importance from an environmental and ecological perspective are some of the motivating factors to study and assess the water quality of this river. This is a real-world use case within the scope of the SWAIN project,[7] which aims to develop an IoT- and AI-driven early warning system for micropollutants in European rivers. One of the main objectives is to achieve continuous monitoring of river water quality at remote places using WSNs with low power consumption, low cost, and high detection accuracy.

We demonstrate the application of our GENS framework to design an efficient, low-cost WQMN and develop a proof-of-concept solution. The network of sensors and gateways is designed to monitor the water quality in Ergene River continuously. Fig. 6 presents the output of the GENS framework and orchestration of a WQMN consisting of the optimal number of sensors and gateways to monitor the Ergene River. The original 75 measurement locations identified by the environmental experts (Fig. 5), which are considered candidate sensor locations in the simulated environment, are shown in Fig. 6(a). It is essential to mention here that these are chosen independently of their ability to transmit the collected data because they were manual sampling locations, and network quality was not taken into account in their selection.

In the next step, sensors are connected to gateways to allow them to transmit the collected data to the monitoring station. The network resulting from the gateway placement module is displayed in Fig. 6(b). Each gateway in this network contains a minimum of two sensors. The network modeling module evaluates the network performance and provides the necessary parameters to calculate the communication cost (CC) associated with each sensor location. The sensor locations that are not included in the candidate network or sensors that are unable to transmit due to connectivity issues are considered high cost and marked as red circles in Fig. 6(b).

Since assembling and maintaining a network of this size is quite challenging, the network proposed by gateway placement is optimized further through the sensor placement module. It keeps sensors that are both informative and communication-efficient and strategically removes sensors whose contributions in reconstructing the original sensor signal are either insignificant or redundant. The WQMN displayed in Fig. 6(c) is used to monitor the water quality of the Ergene River. The WQMN consists of 30 optimal sensors connected through a network of ten gateways. The total number of gateways is reduced from 15 to 10, which is a reduction of 33%, and the number of sensors required to monitor the river is reduced by 40% compared to the original network with 75 sensors.

Ergene River use case also demonstrates that the GENS framework can be used to improve upon an existing monitoring infrastructure. In this circumstances, it would output the minimum cost / maximum information deployment without moving or destroying existing monitoring stations. It is also possible to allow a limited number of changes to further enhance information and communication quality. Considering that conventional WQMN are already available in many rivers, this capability improves the flexibility and practical applicability of the proposed algorithms.

[7] https://swain-project.eu/

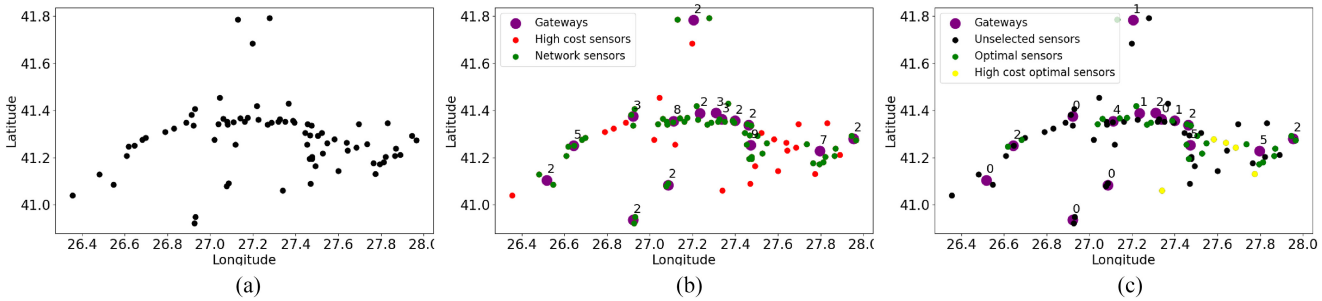

Fig. 6. Monitoring of the Ergene River through optimization of gateways and sensors: (a) original network of 75 sensors, (b) gateway and sensors resulting from the gateway placement module, and (c) set of 30 optimal sensor locations. The numbers above the gateways represent the number of associated sensors.

V. Evaluation

A. Experimental Setup

The experiments are based on a simulation of the above use case scenario, supported with real-world water quality data.

1) Data Set: The data used in this study were collected from 75 locations on the Ergene River between August 2017 and May 2018 using direct injection liquid chromatography-tandem spectrometry. Locations on the river network for assessing the water quality were selected by capturing the micropollutant profile of the major industrial and domestic discharges. Along every major tributary, there exists at least one sampling point. Water samples were collected at each location during four different periods: 1) Summer 2017; 2) Fall 2017; 3) Winter 2018; and 4) Spring 2018. The river flow rate was also measured at each location at the time of sampling. A total of 222 chemical substances and 131 micropollutants in the water samples taken from 75 locations on the Ergene River in four seasons were found in the collected samples [60].

2) Performance Metrics: The evaluation of the proposed framework involves the following five metrics.

CC: Communication is a major energy consumer as the signal power in the sensor network drops as the distance from the transmitter increases. This means that to reach a slightly longer distance, the sensor needs to dispatch much higher transmit power.

We propose two cost metrics to model the CC: 1) network CC (NCC) and 2) gateway cost (GTC). The NCC is defined as the sensor's ability to transfer the data packets successfully. Since each transmission drains the battery, frequent packet failures may result in high retransmissions, which consequently could affect the lifetime of the deployed network. Hence, we have to ensure that our sensor placements have reliable communication links. Sensor networks deployed to monitor outdoor environments can involve sensors that are too far apart. This geographical distance between sensors and gateways can have a significant influence on the overall performance of the network. To model this constraint into our placement algorithm we propose GTC which is defined as the angular (Haversine) distance between the sensors and the closest gateway. The NCC and GTC are defined in (15) and (16), respectively. Here, TPA is the total packet attempts and ST successful transmissions

$$\mathrm{NCC}(s) = \frac{(\mathrm{TPA} - \mathrm{ST})}{\mathrm{TPA}} \tag{15}$$

$$\mathrm{GTC}(s) = 2\arcsin\sqrt{\begin{matrix}\sin^2\left((x_1 - y_1)/2\right) \\ +\cos\left(x_1\right)\cos\left(y_1\right) \\ \sin^2\left((x_2 - y_2)/2\right)\end{matrix}} \tag{16}$$

where the coordinates x and y are for sensors and gateways.

Reconstruction Error: The proposed sensor placement framework is evaluated using the metric called reconstruction error, which as the name suggests measures the sensors ability to reconstruct the area of interest. We use MSE as the metric to estimate the reconstruction error.

Network Performance Value (NPV): In order to evaluate the network simulation of the *Network Modeling Module*, a basic formula called NPV is composed from the individual results

$$\mathrm{NPV} = 100 \times \frac{\mathrm{SS}}{\mathrm{FPA} \times \mathrm{LEC}}. \tag{17}$$

Here, FPA stands for *failed packet attempt* count, LEC *LoRaWAN energy consumption*, and SS *site score* as in

$$\mathrm{SS} = \mathrm{RS} \times (1 - \mathrm{CC}). \tag{18}$$

To quantify the fitness of a particular sensor location, site score takes the reconstruction score (RS) and CC into consideration.

Average Round Trip Time: Round trip time is the time spent per successful packet transmission, which includes transmission of a single packet and receipt of its ACK signal.

Total Energy Consumption: This metric includes the total passive, active, and communication energy consumption of the sensors in kilojoule per hour (kJ/h).

3) Baseline Algorithms: The performance of the MSPQR is compared to the following baseline methods.

Random (RAND): Rand is a baseline algorithm in which sensor locations are deployed randomly in the input space. The system time is used as a seed for generating random locations.

MI-Based (GP-MI): Gaussian process (GP) is used in [34] to place near-optimal sensors. Given a GP model, various criteria have been proposed for assessing the quality of sensor placements, including mutual information (MI) and entropy. GP-MI seeks to maximize the MI between the chosen locations.

Entropy-Based (GP-EN): GP-EN aims to place sensors where uncertainty about the signal is the highest, that is, the highest entropy location of the GP. Placements based on GP-EN result in sensors that are uncertain about each other's measurements [61].

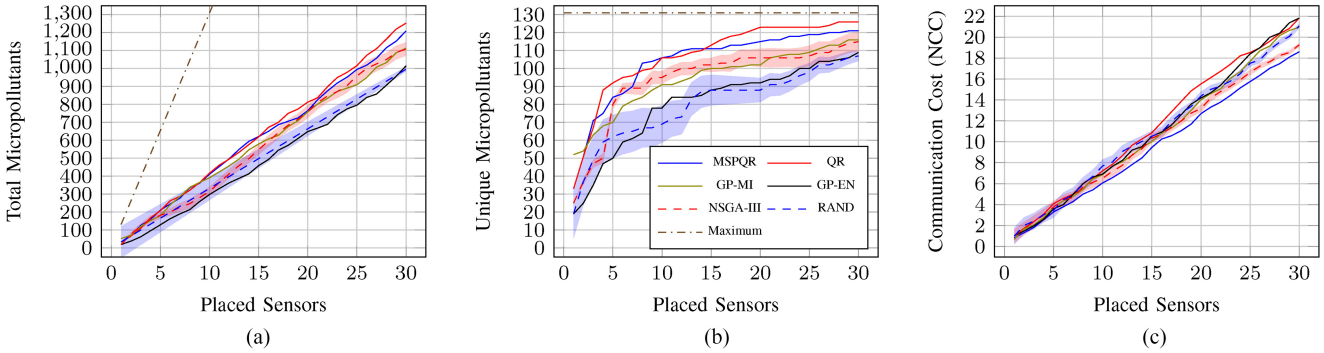

Fig. 7. Evaluation results of MSPQR and baseline algorithms in sensor placement. (a) Total micropollutants detected. (b) Unique micropollutants detected. (c) NCC.

Single-Objective (QR): We implement the single objective QR-based sensor placement (QR) for signal reconstruction [35]. QR aims to find optimal sensor locations without considering any cost constraints. To make the evaluation simple and fair, we use the same test set for assessing the quality of predictions for both MSPQR and QR methods.

Multiobjective (NSGA-III): We implement a nondominated sorting genetic algorithm (NSGA-III) for multiobjective sensor placement as proposed in [62]. We use grid-search to find appropriate values for the hyperparameters, such as population size, crossover, number of offsprings and function evaluations. Moreover, we use pseudo-weights for multicriteria decision-making (MCDM) to choose an optimal solution from the set of nondominated solutions.

For the baselines that include stochastic decisions (RAND and NSGA-III), we provide 95% confidence intervals.

B. Numerical Results

1) Sensor Placement: This section presents the numerical results for reconstructing and predicting the spatiotemporal signal using data collected from the Ergene River.

Fig. 7 compares and contrasts the performance of MSQPR and baseline methods. We compute and compare the number of pollutants detected at the proposed 30 sensor locations for each method, as shown in Fig. 7(a) where the y-axis shows the number of micropollutants detected at each location. The proposed sensor placement method for sensor signal reconstruction takes advantage of the local features of the data and outperforms the baseline methods except for QR in terms of accuracy and the total number of micropollutants detected. The performance of the single-objective QR method is slightly better than MSPQR since QR solely focuses on finding optimal sensor locations without any consideration for the CC or distance to the nearby gateways, which is essential for the sustainable monitoring of large environments such as rivers. The superior performance of QR-based methods over the genetic and GP-based methods is because the tailored bases learned from the local data enable these models to better reconstruct the sensor signal, which in turn helps to find better sensor locations. GP-MI performs comparably well in the earlier phase but is outperformed by both reconstruction-based methods (MSPQR and QR) as the number of sensors increases. Multiobjective NSGA-III does reasonably well and achieves similar results to GP-MI. However, its performance in terms of information quality is not as good as NCC (discussed below). GP-EN is not able to perform as well as GP-MI and is only marginally better than RAND. The lower performance of the GP-EN compared to GP-MI can be explained by the fact that it considers entropy rather than the prediction quality of the selected sensor locations and tends to place sensors at the borders [62]. We also provide the theoretical maximum value calculated as (number of sensors) (131 unique micropollutants). This limit is not generally reached, as not all micropollutants are present in every part of the river.

Results of unique micropollutants detected, displayed in Fig. 7(b), are generally consistent with Fig. 7(a). The high number of unique micropollutants detected by MSPQR and QR (121 and 126 of maximum 131) highlights the fact that reconstruction-based models can better capture the variance of the original sensor signal than the other baseline methods. For instance, two sensors that happen to be on the same tributary but at different altitudes might be measuring redundant information. NSGA-III and GP-based methods are not able to discern between the two as (mutual) information provided by these sensors is likely to be similar. Therefore, both sensors would be regarded as optimal locations. On the contrary, any sensor which does not contribute to the reconstruction of the original signal or measures redundant information would be given low scores and hence removed from the placements.

Fig. 7(c) illustrates how CC changes with respect to the number of sensors. The performance of MSPQR and NSGA-III methods is similar up to a total of 15 sensors. However, as the number of sensors increases, NSGA-III performance deteriorates a little, while MSPQR consistently outperforms NSGA-III. It is worth mentioning that during the selection of the optimal solution from the Pareto front, we assigned a slightly higher weight to the CC objective. This allows NSGA-III to achieve comparable performance to MSPQR but at the expense of information quality. Regarding the other baselines, their performance is similar up to ten sensors. However, as the number of sensors increases, the nonconsideration of CC leads to an increase in NCC. This indicates that MSPQR aims to find the balance between the reconstruction quality and CC whereas the other baselines, except NSGA-III, are solely concerned with the information

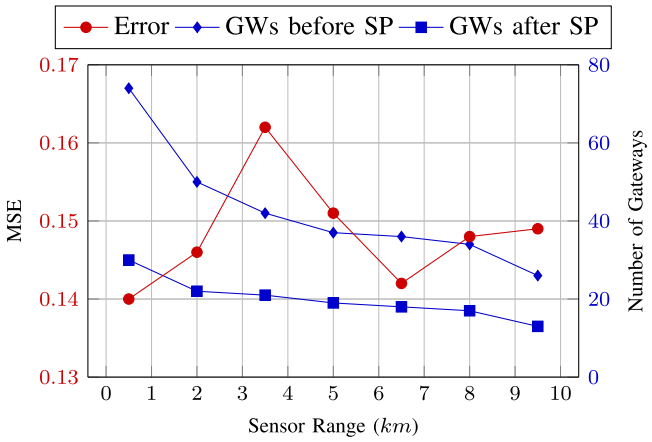

Fig. 8. Placement results with an increasing sensor range.

quality resulting in higher CCs. It is clear that MSPQR is able to better identify sensor locations that are both information and communication efficient compared to the baselines.

Fig. 8 summarizes the performance of MSPQR and the impact of sensor range on reconstruction error. The number of gateways and sensing range affects the CC, which consequently influences the models' reconstruction ability. Therefore, we performed a series of experiments varying the number of gateways and sensing range, and the obtained results are presented in Fig. 8. The results shown in this figure were obtained from an average of ten trials. The sensor placement model is able to reconstruct the spatiotemporal signal given the limited number of measurements. The sharp rise in reconstruction error, especially when the sensor's range is relatively shorter (between 2 and 3.5 km), may seem counterintuitive. However, when combined with the fact that the network resulting from this configuration involves the most high-cost sensors, 31 and 28 for the 2 and 3.5-km range, respectively. When the network contains more high-cost sensors, the selected sensors are pushed out of the regions of high cost. This allows the model to keep the CC low at the expense of increased reconstruction error. The method achieves the best reconstruction accuracy for spatiotemporal signal reconstruction for configuration in which the sensor range is set to 6.5 km. Overall, the network resulting after the sensor placement contains 33%–50% fewer gateways.

As mentioned briefly in Section III-C, the type of cost used to characterize the cost associated with each sensor location can degrade or enhance the performance of the deployed network. Depending upon the application and size of the area to be monitored, different cost metrics could yield superior results. In order to study the influence of the type of cost on network performance and to identify the settings in which NCC yields better results than GTC and vice versa, we perform some experiments and report the results in Fig. 9. In network configurations with sensors having a limited range and, thus, being closer to the gateways, GTC outperforms NCC and results in sensor locations that are more informative about the water quality as shown in Fig. 9(a). Networks involving sensors within the sensing range between 500–4000 m detect a higher number of micropollutants with GTC than NCC. However, when the cost function weight is increased (which means the distance between the sensors and gateways becomes larger), sensors are gradually pushed out of the regions of high cost. This allows the total cost to be lowered at the expense of reducing reconstruction accuracy [Fig. 9(b)]. NCC, on the other hand, is oblivious to the distance between the sensors and gateways and is only concerned with the sensor's ability to transmit data successfully. Therefore, the network's performance involving high-range sensors is not affected as much as it is for GTC. As evident from both Fig. 9(a) and (b), NCC is able to achieve similar or better performance when the sensor range is between 6500–9500 m than GTC.

2) Gateway Placement: Fig. 10(a) shows the NPV results, calculated from the simulation of 30 sensors which were assigned the highest scores by the sensor placement module. The number of sensors is progressively reduced in order to observe the performance effect of taking as few sensors as possible. Results are also obtained for different signal ranges to further observe the effect of distance on the performance. The simulation environment follows the EU regulations, which are determined by the European Telecommunications Standards Institute (ETSI) for the LoRaWAN duty-cycle limitations for the 868.0–868.6-MHz sub-bands and sets the duty-cycle limit as 1% [63], [64]. Since sensor communication will rely on a private ad-hoc network, we do not consider fair use policies. From the results, it can be seen that when the signal range is 1 km, all of the simulation results show a linear drop in performance as the number of sensors is decreased. However, as the signal range increases, while the overall performance drops for all scenarios, the best results are achieved by picking the top ten sensors. The reason is that extending signal range causes more interference between packets transmission with a consequent increase in retransmission attempts. Moreover, additional sensors also cause interferences, so performance degradation becomes more evident with larger sensor counts.

Another result seen in Fig. 10(b) is the average time spent per successful packet transmission. Intuitively, when the signal range gets shorter, the time spent to transmit a packet and receive its ACK also decreases. Following from NPV result, when the distance is 1 km, all of the sensors nearly spend the same amount of time for the transmission process. However, as the signal distance increases, using a high number of sensors extend the successful packet transmission time since failure repetitions and duty-cycle limitations affect the performance.

3) Energy Optimization: We compare energy optimization (GENETIC) with the baselines analyzed in previous phases, namely, RAND, MSQPR, QR, GPMI, GP-EN, and NSGA-III. Also, we present the result of $E(P)$ for MAX-NP, i.e., the placement allowing detection of most micropollutants computed in the previous GENS phases, to demonstrate that GENS is capable of achieving energy savings without losing its sensing capability. Results in Fig. 11 show energy consumption per successful transmission in Joules. We see that energy optimization allows for reducing energy consumption up to 60% without affecting the number of micropollutants detected. We also observe that it outperforms the best performing baseline in terms of energy consumption, MSPQR, by around 10%. Results indicate that by just tuning gateway ranges, we can

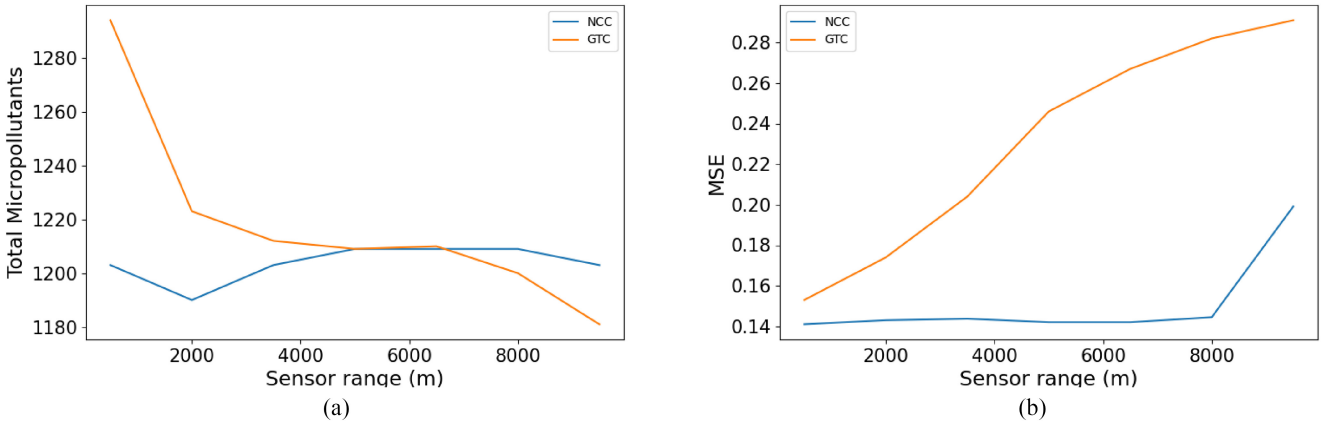

Fig. 9. Comparison between NCC and GTC. (a) Impact on number of MPs detected. (b) Impact on MSE.

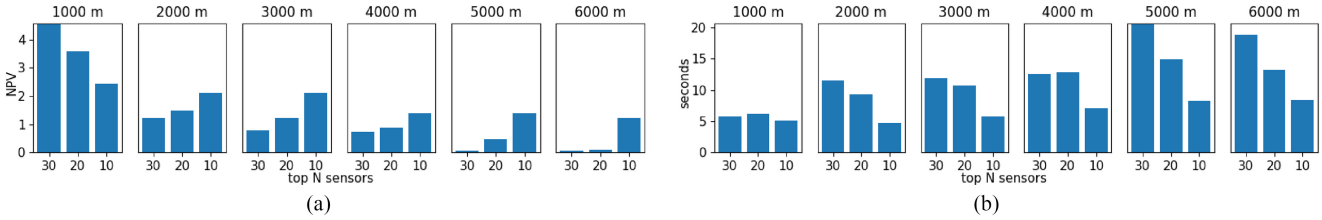

Fig. 10. Results obtained from simulating the designed sensor-gateway topology in NS-3 network simulator for top N sensors. (a) NPV. (b) Average round trip time.

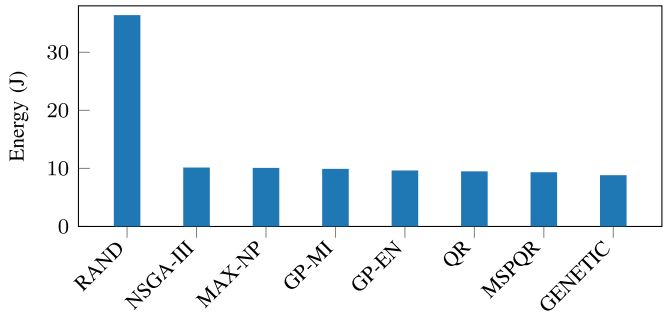

Fig. 11. Average $E(P)$ per successful transmission.

achieve substantial energy improvement, which is of capital importance in low-energy scenarios where GENS operates.

VI. Conclusion

In this work, we study the multinode sensor placement problem, i.e., given a set of probable sensor locations, determine a set of sensor locations that are both information and communication efficient. Once optimal sensor locations are found, place a minimum number of gateway sensors to maintain the global connectivity such that the resulting network of sensors is low cost and energy efficient. Our study is motivated by an important class of WSNs, in which IoT nodes that are responsible for measuring the environmental parameters (e.g., pH, turbidity, etc., in the case of water quality monitoring) required to monitor the environment are connected with the gateway through LoRaWAN technology in order to transfer the collected data to a base station or a processing center to store or further process the data.

We propose the GENS framework, which is comprised of a simulation-based gateway placement module, a QR decomposition-based sensor placement module, and a genetic algorithm-based energy optimization module. Extensive experiments subject to real-world water quality data from Ergene River, Turkey, demonstrate superior performance compared to state-of-the-art baselines in terms of information quality, communication capacity, and energy efficiency. The contributions in this work will be implemented in practice within the scope of the SWAIN project, which aims to develop an IoT- and AI-driven early warning system for pollutants in European rivers.

References

[1] UN General Assembly. "Transforming our world: The 2030 agenda for sustainable development." 2015. [Online]. Available: https://undocs.org/en/A/RES/70/1

[2] A. Fascista, "Toward integrated large-scale environmental monitoring using WSN/UAV/crowdsensing: A review of applications, signal processing, and future perspectives," *Sensors*, vol. 22, no. 5, p. 1824, 2022.

[3] G. Xu, W. Shen, and X. Wang, "Applications of wireless sensor networks in marine environment monitoring: A survey," *Sensors*, vol. 14, no. 9, pp. 16932–16954, 2014.

[4] M. Z. A. Bhuiyan, G. Wang, J. Cao, and J. Wu, "Sensor placement with multiple objectives for structural health monitoring," *ACM Trans. Sens. Netw.*, vol. 10, no. 4, pp. 1–45, 2014.

[5] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.

[6] K. Fukami, R. Maulik, N. Ramachandra, K. Fukagata, and K. Taira, "Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning," *Nat. Mach. Intell.*, vol. 3, no. 11, pp. 945–951, 2021.

[7] S. L. Ullo and G. R. Sinha, "Advances in smart environment monitoring systems using IoT and sensors," *Sensors*, vol. 20, no. 11, p. 3113, 2020.

[8] T. Inoue et al., "Data-driven optimal sensor placement for high-dimensional system using annealing machine," *Mech. Syst. Signal Process.*, vol. 188, Apr. 2023, Art. no. 109957.

[9] L. M. Oliveira and J. J. Rodrigues, "Wireless sensor networks: A survey on environmental monitoring," *J. Commun.*, vol. 6, no. 2, pp. 143–151, 2011.

[10] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg, "Near-optimal sensor placements: Maximizing information while minimizing communication cost," in *Proc. Int. Conf. Inf. Process. Sens. Netw.*, 2006, pp. 2–10.

[11] M. A. Benatia, M. Sahnoun, D. Baudry, A. Louis, A. El-Hami, and B. Mazari, "Multi-objective WSN deployment using genetic algorithms under cost, coverage, and connectivity constraints," *Wireless Pers. Commun.*, vol. 94, no. 4, pp. 2739–2768, 2017.

[12] X. Wu, G. Chen, and S. K. Das, "On the energy hole problem of nonuniform node distribution in wireless sensor networks," in *Proc. IEEE Int. Conf. Mobile Ad Hoc Sens. Syst.*, 2006, pp. 180–187.

[13] T. Ojha, S. Misra, and N. S. Raghuwanshi, "Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges," *Comput. Electron. Agr.*, vol. 118, pp. 66–84, Oct. 2015.

[14] H. M. Jawad, R. Nordin, S. K. Gharghan, A. M. Jawad, and M. Ismail, "Energy-efficient wireless sensor networks for precision agriculture: A review," *Sensors*, vol. 17, no. 8, p. 1781, 2017.

[15] R. Altenburger et al., "Future water quality monitoring—Adapting tools to deal with mixtures of pollutants in water resource management," *Sci. Total Environ.*, vols. 512–513, pp. 540–551, Apr. 2015.

[16] S. K. Ning and N.-B. Chang, "Multi-objective, decision-based assessment of a water quality monitoring network in a river system," *J. Environ. Monit.*, vol. 4, no. 1, pp. 121–126, 2002.

[17] Q. Chen, W. Wu, K. Blanckaert, J. Ma, and G. Huang, "Optimization of water quality monitoring network in a large river by combining measurements, a numerical model and matter-element analyses," *J. Environ. Manage.*, vol. 110, pp. 116–124, Nov. 2012.

[18] C. M. G. Carpenter and D. E. Helbling, "Widespread micropollutant monitoring in the hudson river estuary reveals spatiotemporal micropollutant clusters and their sources," *Environ. Sci. Technol.*, vol. 52, no. 11, pp. 6187–6196, 2018.

[19] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, "Long-range communications in unlicensed bands: The rising stars in the IoT and smart city scenarios," *IEEE Wireless Commun.*, vol. 23, no. 5, pp. 60–67, Oct. 2016.

[20] K. Mekki, E. Bajic, F. Chaxel, and F. Meyer, "Overview of cellular LPWAN technologies for IoT deployment: Sigfox, LoRaWAN, and NB-IoT," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, 2018, pp. 197–202.

[21] P. Ruckebusch, S. Giannoulis, I. Moerman, J. Hoebeke, and E. De Poorter, "Modelling the energy consumption for over-the-air software updates in LPWAN networks: SigFox, LoRa and IEEE 802.15.4G," *Internet Things*, vols. 3–4, pp. 104–119, Oct. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2542660518300362

[22] M. Iqbal, A. Y. M. Abdullah, and F. Shabnam, "An application based comparative study of LPWAN technologies for IoT environment," in *Proc. IEEE Region 10 Symp. (TENSYMP)*, 2020, pp. 1857–1860.

[23] H. Mroue, A. Nasser, S. Hamrioui, B. Parrein, E. Motta-Cruz, and G. Rouyer, "MAC layer-based evaluation of IoT technologies: LoRa, SigFox and NB-IoT," in *Proc. IEEE Middle East North Afr. Commun. Conf. (MENACOMM)*, 2018, pp. 1–5.

[24] A. Lavric and V. Popa, "A LoRaWAN: Long range wide area networks study," in *Proc. Int. Conf. Electromech. Power Syst. (SIELMEN)*, 2017, pp. 417–420.

[25] D. Carrillo and J. Seki, "Rural area deployment of Internet of Things connectivity," in *Proc. IEEE Int. Conf. Electron., Electr. Eng. Comput. (INTERCON)*, 2017, pp. 1–4.

[26] A. Mdhaffar, T. Chaari, K. Larbi, M. Jmaiel, and B. Freisleben, "IoT-based health monitoring via LoRaWAN," in *Proc. IEEE Int. Conf. Smart Technol.*, 2017, pp. 519–524.

[27] M. Aernouts, R. Berkvens, K. Van Vlaenderen, and M. Weyn, "Sigfox and LoRaWAN datasets for fingerprint localization in large urban and rural areas," *Data*, vol. 3, no. 2, p. 13, 2018.

[28] A. Grunwald, M. Schaarschmidt, and C. Westerkamp, "LoRaWAN in a rural context: Use cases and opportunities for agricultural businesses," in *Proc. Mobile Commun. Technol. Appl. ITG-Symp.*, 2019, pp. 1–6.

[29] B. Ousat and M. Ghaderi, "LoRa network planning: Gateway placement and device configuration," in *Proc. IEEE Int. Congr. Internet Things (ICIOT)*, 2019, pp. 25–32.

[30] S. Mnguni, A. M. Abu-Mahfouz, P. Mudali, and M. O. Adigun, "A review of gateway placement algorithms on Internet of Things," in *Proc. Int. Conf. Adv. Big Data, Comput. Data Commun. Syst. (icABCD)*, 2019, pp. 1–6.

[31] N. Matni, J. Moraes, H. Oliveira, D. Rosário, and E. Cerqueira, "LoRaWAN gateway placement model for dynamic Internet of Things scenarios," *Sensors*, vol. 20, no. 15, p. 4336, 2020.

[32] F. Loh, N. Mehling, S. Geißler, and T. Hoßfeld, "Graph-based gateway placement for better performance in LoRaWAN deployments," in *Proc. 20th Mediterr. Commun. Comput. Netw. Conf. (MedComNet)*, 2022, pp. 190–199.

[33] F. Loh, D. Bau, J. Zink, A. Wolff, and T. Hoßfeld, "Robust gateway placement for scalable LoRaWAN," in *Proc. 13th IFIP Wireless Mobile Netw. Conf. (WMNC)*, 2021, pp. 71–78.

[34] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, no. 2, pp. 235–284, 2008.

[35] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, "Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns," *IEEE Control Syst. Mag.*, vol. 38, no. 3, pp. 63–86, Jun. 2018.

[36] Z. Drmac and S. Gugercin, "A new selection operator for the discrete empirical interpolation method—Improved a priori error bound and extensions," *SIAM J. Sci. Comput.*, vol. 38, no. 2, pp. A631–A648, 2016.

[37] Y. Guo, Y. Ni, and S. Chen, "Optimal sensor placement for damage detection of bridges subject to ship collision," *Struct. Control Health Monit.*, vol. 24, no. 9, 2017, Art. no. e1963.

[38] Z. Xu, Y. Guo, and J. H. Saleh, "Multi-objective optimization for sensor placement: An integrated combinatorial approach with reduced order model and Gaussian process," *Measurement*, vol. 187, 2022, Art. no. 110370.

[39] R. Hou, Y. Xia, Q. Xia, and X. Zhou, "Genetic algorithm based optimal sensor placement for L1-regularized damage detection," *Struct. Control Health Monit.*, vol. 26, no. 1, 2019, Art. no. e2274.

[40] B. Isong, R. R. S. Molose, A. M. Abu-Mahfouz, and N. Dladlu, "Comprehensive review of SDN controller placement strategies," *IEEE Access*, vol. 8, pp. 170070–170092, 2020.

[41] T. Kim, B. D. Youn, and H. Oh, "Development of a stochastic effective independence (SEFI) method for optimal sensor placement under uncertainty," *Mech. Syst. Signal Process.*, vol. 111, pp. 615–627, Oct. 2018.

[42] B. Kabakulak, "Sensor and sink placement, scheduling and routing algorithms for connected coverage of wireless sensor networks," *Ad Hoc Netw.*, vol. 86, pp. 83–102, Apr. 2019.

[43] F. A. Hashim, E. H. Houssein, M. S. Mabrouk, W. Al-Atabany, and S. Mirjalili, "Henry gas solubility optimization: A novel physics-based algorithm," *Future Gener. Comput. Syst.*, vol. 101, pp. 646–667, Dec. 2019.

[44] J. Wang, J. Cao, R. S. Sherratt, and J. H. Park, "An improved ant colony optimization-based approach with mobile sink for wireless sensor networks," *J. Supercomput.*, vol. 74, no. 12, pp. 6633–6645, 2018.

[45] H. Banka and P. K. Jana, "PSO-based multiple-sink placement algorithm for protracting the lifetime of wireless sensor networks," in *Proc. 2nd Int. Conf. Comput. Commun. Technol. IC3T*, vol. 1, 2016, pp. 605–616.

[46] S. Mohammadi and G. Farahani, "A new algorithm inspired by impala Mexican wave with variable stride for relay node placement as a nested reverse p-median problem in disjoint wireless sensor networks," *Wireless Netw.*, vol. 27, no. 4, pp. 2347–2363, 2021.

[47] A. Jari and A. Avokh, "PSO-based sink placement and load-balanced anycast routing in multi-sink WSNs considering compressive sensing theory," *Eng. Appl. Artif. Intell.*, vol. 100, Apr. 2021, Art. no. 104164.

[48] C. Zhao et al., "Maximizing lifetime of a wireless sensor network via joint optimizing sink placement and sensor-to-sink routing," *Appl. Math. Model.*, vol. 49, pp. 319–337, Sep. 2017.

[49] A. Shamshirgaran, H. Javidi, and D. Simon, "Evolutionary algorithms for multi-objective optimization of drone controller parameters," in *Proc. IEEE Conf. Control Technol. Appl. (CCTA)*, 2021, pp. 1049–1055.

[50] A. Aral, V. De Maio, and I. Brandic, "ARES: Reliable and sustainable edge provisioning for wireless sensor networks," *IEEE Trans. Sustain. Comput.*, vol. 7, no. 4, pp. 761–773, Oct.–Dec. 2022.

[51] G. F. Riley and T. R. Henderson, "The NS-3 network simulator," in *Modeling and Tools for Network Simulation*, K. Wehrle, M. Güneş, and J. Gross, Eds. Heidelberg, Germany: Springer, 2010, pp. 15–34.

[52] D. Magrin, M. Centenaro, and L. Vangelista, "Performance evaluation of LoRa networks in a smart city scenario," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2017, pp. 1–7.

[53] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, and T. Watteyne, "Understanding the limits of LoRaWAN," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 34–40, Sep. 2017.
[54] S. Phaiboon and S. Somkuarnpanit, "Mobile path loss characteristics for low base station antenna height in different forest densities," in *Proc. Int. Symp. Wireless Pervasive Comput.*, 2006, p. 6.
[55] E. Clark, T. Askham, S. L. Brunton, and J. N. Kutz, "Greedy sensor placement with cost constraints," *IEEE Sensors J.*, vol. 19, no. 7, pp. 2642–2656, Apr. 2019.
[56] H. Cheng, Z. Gimbutas, P.-G. Martinsson, and V. Rokhlin, "On the compression of low rank matrices," *SIAM J. Sci. Comput.*, vol. 26, no. 4, pp. 1389–1404, 2005.
[57] M. Gu and S. C. Eisenstat, "Efficient algorithms for computing a strong rank-revealing QR factorization," *SIAM J. Sci. Comput.*, vol. 17, no. 4, pp. 848–869, 1996.
[58] C. Li, S. Jegelka, and S. Sra, "Polynomial time algorithms for dual volume sampling," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.
[59] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: Past, present, and future," *Multimedia Tools Appl.*, vol. 80, pp. 8091–8126, Feb. 2021.
[60] S. M. Emadian, F. O. Sefiloglu, I. A. Balcioglu, and U. Tezel, "Identification of core micropollutants of Ergene river and their categorization based on spatiotemporal distribution," *Sci. Total Environ.*, vol. 758, Mar. 2021, Art. no. 143656.
[61] N. Ramakrishnan, C. Bailey-Kellogg, S. Tadepalli, and V. N. Pandey, "Gaussian processes for active data mining of spatial aggregates," in *Proc. SIAM Int. Conf. Data Min.*, 2005, pp. 427–438.
[62] C. Hu, L. Dai, X. Yan, W. Gong, X. Liu, and L. Wang, "Modified NSGA-III for sensor placement in water distribution system," *Inf. Sci.*, vol. 509, pp. 488–500, Jan. 2020.
[63] *Short Range Devicesoperating in the Frequency Range 25 MHz to 1000 MHz; Part 1: Technical Characteristics and Methods of Measurement*, ETSI Standard EN 300 220-1, V3.1.1, Feb. 2017.
[64] "Technical characteristics for low power wide area networks chirp spread spectrum (LPWAN-CSS) operating in the UHF spectrum below 1 GHz," Eur. Telecommun. Stand. Inst., Sophia Antipolis, France, Rep. TR 103 526, V1.1.1, Apr. 2018.

Sabtain Ahmad (Graduate Student Member, IEEE) received the B.Sc. degree in computer science from the National University of Computer and Emerging Sciences, Lahore, Pakistan, in 2015, and the M.Sc. degree in computer science from the Technical University of Berlin, Berlin, Germany, in 2020. He is currently pursuing the Ph.D. degree with Vienna University of Technology, Vienna, Austria.

His research interests include artificial intelligence (AI) for resource-constrained environments (Edge-AI) and privacy-preserving distributed machine learning.

Halit Uyanık received the B.Sc. and M.Sc. degrees in computer engineering from Istanbul Technical University, Istanbul, Turkey, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree.

His research interests include consistency and resource management in distributed systems, software quality, and performance management of IoT devices.

Tolga Ovatman (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in computer engineering from Istanbul Technical University (İTÜ), Istanbul, Turkey, in 2005 and 2011, respectively.

He is working as an Associate Professor with the Computer Engineering Department, İTÜ. His research interests include cloud computing, object-oriented software analysis, and design and formal verification.

Mehmet Tahir Sandıkkaya (Senior Member, IEEE) received the B.S. degree in electrical engineering from Istanbul Technical University (ITU), Istanbul, Turkey, in 2002, and the M.S. and Ph.D. degrees in computer science from Informatics Institute, ITU in 2005 and 2015, respectively.

He currently holds an Assistant Professorship position with the Computer Engineering Department, ITU. His research area is computer security and privacy, where his recent activities focus on cloud computing security, security measurements, security protocols, and cyber–physical system's security.

Vincenzo De Maio (Member, IEEE) received the Ph.D. degree from the University of Innsbruck, Innsbruck, Austria, in 2016.

His research in the area of parallel and distributed systems comprises energy-aware cloud/edge computing and scheduling. Since 2017, he has been a Postdoctoral Researcher with Vienna University of Technology, Vienna, Austria. His main research interests include cloud/edge placement, energy efficiency in distributed systems, sustainable computing, and quantum computing.

Ivona Brandić (Member, IEEE) received the Ph.D. degree from Vienna University of Technology, Vienna, Austria, in 2007.

She is a Full Professor with Vienna University of Technology. Her main research interests are cloud computing, large-scale distributed systems, energy efficiency, QoS, and autonomic computing.

Prof. Brandić was awarded the FWF START Prize, the highest Austrian award for early career researchers in 2015. In 2011, she received the Distinguished Young Scientist Award from Vienna University of Technology for her project on Holistic Energy-Efficient Hybrid Clouds.

Atakan Aral (Senior Member, IEEE) received the dual M.Sc. degree in computer science and engineering from Politecnico di Milano, Milan, Italy, in 2011, and Istanbul Technical University (ITU), Istanbul, Turkey, in 2012, and the Ph.D. degree in computer engineering from ITU in 2016.

He is a Tenure-Track Assistant Professor with the Department of Computing Science, Umeå University, Umeå, Sweden, and a Research Fellow with the Faculty of Computer Science, University of Vienna, Vienna, Austria. His research interests center around resource management for geo-distributed and virtualized systems, such as intercloud and edge computing.