Go with the Flow: Leveraging Physics-Informed Gradients to Solve Real-World Problems in Water Distribution Systems

Inaam Ashraf[1,2], Janine Strotherm[1,2](✉), Luca Hermes[1,2], and Barbara Hammer[2]

[1] Authors contributed equally
[2] Faculty of Technology, Bielefeld University, Inspiration 1, 33619 Bielefeld
{mashraf,jstrotherm,lhermes,bhammer}@techfak.uni-bielefeld.de

Abstract. Clean drinking water is essential for a sustainable society as emphasized by UN's sustainable developmental goal 6. Efficient management of water distribution systems (WDSs) is vital to ensure this goal. Conventional approaches rely on computationally expensive hydraulic simulations. Instead, using a pre-trained physics-informed graph neural network as a surrogate model, we solve such real-world problems with gradient methods. This does not only enable end-to-end optimization of WDS attributes but demonstrates the more general concept of leveraging the differentiability of a deep surrogate model to solve downstream tasks related to the underlying complex system. In this work, we demonstrate this novel principle by focusing on three tasks: First, we estimate hydraulic states from sparse sensory information, achieving SOTA performance. Second, we use the surrogate model combined with information theory to solve the task of optimal sensor placement. We use the sparse-to-dense pressure estimation task to gauge the quality of our sensor placements, which itself is non-trivial. Finally, we plan the rehabilitation of WDSs by optimizing pipe diameters in response to changing demands. To the best of our knowledge, we are the first to use the concept of end-to-end differentiability of complex systems via deep surrogate models to solve real-world tasks in WDSs.

Keywords: Physics-informed Machine Learning · Graph neural networks · Surrogate models · Digital twins · Water distribution networks

1 Introduction

The increasing availability of physics-informed machine learning (ML) methodologies allows the development of efficient and reliable surrogate models, i.e., ML models that mimic the physics of the real world [16]. Integration of such surrogate models into real-world scenarios makes them *digital twins* [13]. Critical infrastructure, such as energy and transportation systems as well as water distribution systems (WDSs), is one example where these models are of high relevance [4]. Here, important functionality ranges from reliable state estimation, predictive control, up to improved resilience.

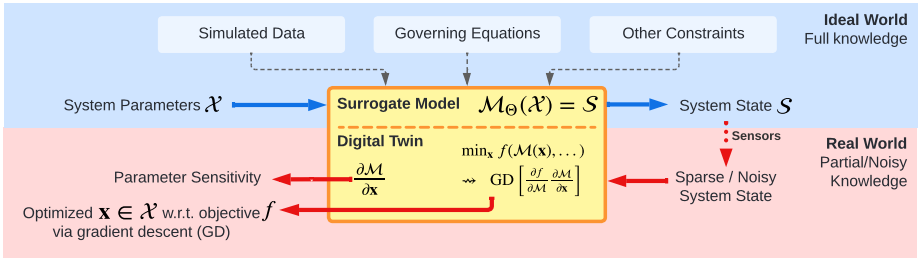


Fig. 1: The main concept of leveraging gradients of a surrogate model $\mathcal{M}$ with parameters Θ. Given a surrogate model (top) that predicts system states $\mathcal{S}$ from system parameters $\mathcal{X}$, its derivative can be used to solve real-world problems as a digital twin (bottom) like sensitivity analysis, inferring the system state, or optimizing system parameters $\mathbf{x} \in \mathcal{X}$ given an objective f. This only requires noisy/sparse state information.

We focus on one specific example, namely WDSs, where both data and models are available. In this context, states such as demands, pressures and flows are of high importance, but are also subject to high uncertainty. A state of the art (SOTA) surrogate model for WDSs has been proposed recently in the form of a physics-informed graph convolutional network (PI-GCN), which is trained based on realistic demands and underlying hydraulic principles [2].

However, while surrogate models typically aim to learn a physically well-defined dynamic directly, tasks in real-world settings often depend on partially unknown quantities. This results in optimization targets that are ill-posed or require architectures which are narrowly tailored to the problem at hand. We want to leverage the property of surrogate models to intrinsically represent the underlying physical structure. Figure 1 highlights our general idea that the backward pass of surrogates can be used to align any combination of known and unknown variables in a physically meaningful way and hence leverage this information to an efficient solution of downstream tasks which occur in this context. Specifically, due to fast computation and the differentiability that comes with a deep learning (DL) approach, such surrogate models enable us to solve important water-related downstream tasks in an end-to-end manner.

In this work, we will formalize the general principle and demonstrate its potential on three relevant tasks: Sparse-to-dense pressure estimation, sensor placement, and network rehabilitation. Although applied in a domain-specific context, this work targets a broader picture, as differentiability allows us to optimize any variable associated with the system that is an input to the surrogate model. This opens the floor to further downstream tasks in real-world scenarios, yielding flexible digital twins based on the backward pass of the surrogate model.

Contributions. Our contributions are threefold:

1. We model the general principle of using the backward pass of a physics-informed surrogate model as a digital twin, which allows to efficiently solve downstream tasks related to the underlying system.
2. We showcase the practicality of this concept in the context of WDSs:
   (a) We propose a new approach for the sparse-to-dense pressure estimation task with high generalizability. As this problem is ill-posed, our method generates a distribution of physically plausible hydraulic states.
   (b) We propose an improved approach to sensor placement by using the gradients of a PI-GCN surrogate model as a measure of sensitivity in combination with information theory.
   (c) We propose a new approach for the network rehabilitation task by gradient-based optimization of the diameters of the pipes that are inputs to the PI-GCN surrogate model.
3. We compare all methods to the SOTA baselines and thoroughly justify our evaluation methods.

Social and domain impact. Since WDSs belong to critical infrastructure, every methodology that improves the SOTA, for example in terms of accuracy or computational complexity, has a high social impact as it improves urban monitoring, planning and rehabilitation. Our proposed methods do not only satisfy this criterion but are also immediately applicable and extendable to other systems of critical infrastructure, given the availability of a differentiable surrogate model of the system.

2 Related Work

State estimation in WDSs. A WDS consists of nodes such as junctions, reservoirs, and tanks connected via links such as pipes, pumps, and valves. Its physical states are characterized by pressure heads and demands at the nodes and flows through links. State estimation describes the task of mapping nodal demands and reservoir pressure heads to flows and pressure heads in the whole WDS. This task can be solved with hydraulic simulators such as EPANET [17]. ML surrogate models have been proposed to solve the state estimation task as a replacement for EPANET [2,12,23]. The most recent SOTA in DL is a PI-GCN [2]. This model is the only one that is regularized by the physical properties of the underlying system; it is less prone to overfit the training distribution.

Sparse-to-dense pressure estimation. For downstream tasks, the pressure heads at every node in a WDS are of interest. In the real world, however, pressure is only known at locations where sensors are installed. Simulators such as EPANET could be used to generate the dense pressure state from demands, but the demands themselves are fully or partially unknown. Hence, simulators or surrogates alone cannot solve this problem. Estimating the full state from sparse sensor readings – sparse-to-dense state estimation – constitutes an ill-posed problem

since multiple states can correspond to the same sensor readings; no analytic solutions have been proposed for this task. Recent DL approaches tackle this problem based on intrinsic regularization [1,10,20]. All of these works employ different types of graph convolutional networks (GCNs) with good results, but limited generalization to unseen WDS attributes.

Sensor placement. WDS operations depend on information from sensors. As providing each node with a sensor is infeasible, sensor placement at selected junctions is a well-known optimization task [5,9,11,14,18]. Some existing algorithms come from the water domain directly [9,18]; recently, ML methods were leveraged [5,14]. Our work combines and improves concepts from [9] and [18] to solve the task of sensor placement. [9] solve a multi-objective optimization problem (OP) based on pressure sensitivity, estimated from hand-crafted gradients, and entropy, derived from these gradients as a measure of network coverage. We access the actual gradients of the differentiable PI-GCN surrogate model instead and use mutual information that assures potential sensors with high pressure sensitivity not to be redundant. [18] also use mutual information, but depend on the downstream task of leakage detection.

Network rehabilitation. WDS planning and rehabilitation is a complicated and multi-objective task [15,21]. Given constraints and uncertainties in the future, planning and rehabilitation tasks are usually formulated as a multi-objective OP [3,6]. In the water community, genetic algorithms such as NSGA-II [7] are used to solve the latter, which are computationally expensive.

3 Background: Tasks in Water Distribution Systems

A WDS can be modeled as a graph, consisting of nodes $V = \{v_1, \dots, v_{n_\mathrm{n}}\}$ (consumer junctions, reservoirs and tanks) and edges $E = \{e_1, \dots, e_{n_\mathrm{e}}\} = \{e_{vu} \mid \forall v \in V, u \in \mathcal{N}(v)\}$ (pipes, pumps and valves). Among all nodes, we consider the set of reservoir nodes V_r, the consumer nodes $V_\mathrm{c} = V \setminus V_\mathrm{r}$ and the sensor nodes $V_\mathrm{s} \subset V_\mathrm{c}$, which are equipped with a sensor that measures the pressure head $h_v \in \mathbb{R}_+$ at the corresponding node $v \in V_\mathrm{s}$. The state of a WDS is characterized by pressure heads $\mathbf{h} = (h_v)_{v \in V} \in \mathbb{R}_+^{n_\mathrm{n}}$ at every node and water demands $\mathbf{d} = (d_v)_{v \in V_\mathrm{c}} \in \mathbb{R}_+^{n_\mathrm{c}}$ at every consumer node along with the water flows $\mathbf{q} = (q_e)_{e \in E} \in \mathbb{R}^{n_\mathrm{e}}$ through every pipe. The relationships between these variables are governed by hydraulic principles, summarized in appendix A.1. A triplet $(\mathbf{h}, \mathbf{d}, \mathbf{q})$ that satisfies these hydraulics is called *physically correct*.

State Estimation. An important task in WDS is to estimate physically correct states $(\mathbf{h}, \mathbf{d}, \mathbf{q})$ of a WDS given initial conditions, like pressure heads $\mathbf{h}_{V_\mathrm{r}} := (h_v)_{v \in V_\mathrm{r}}$ at the reservoirs and the demands $\mathbf{d} = (d_v)_{v \in V_\mathrm{c}}$ at the consumer nodes. We will stick to the convention of subscripting a set to a vector if the vector is limited to that set. Especially, $\mathbf{h}_V = \mathbf{h}, \mathbf{d}_{V_\mathrm{c}} = \mathbf{d}$ and $\mathbf{q}_E = \mathbf{q}$ holds. The first DL approach [2] for state estimation based on reservoir heads and consumer demands

only is fully differentiable, unlike the SOTA hydraulic simulator EPANET [17]. This enables us to propose novel solutions for typical challenges in WDSs, which we present in the next subsections.

3.1 Sparse to Dense Pressure Estimation

Definition 1 (Sparse-to-dense pressure estimation). *Given a set of reservoir and sensor nodes V_r and V_s in a WDS, and observed pressure heads $\mathbf{h}_{V_r \cup V_s}$, sparse-to-dense pressure estimation aims at estimating the pressure heads $\mathbf{h}$ in the whole network.*

Baselines. This task was first approached by [1,10]. [10] train a spectral GCN that requires sparse heads $\mathbf{h}_{V_r \cup V_s}$ as inputs to obtain the heads $\mathbf{h}$ as an output. The more recent work of [1] presents an improved but less generalizable GCN that also requires pipe attributes in addition to the sparse pressures $\mathbf{h}_{V_r \cup V_s}$.

3.2 Sensor Placement

Definition 2 (Sensor placement). *Given a set of consumer nodes $V_c \subset V$ in a WDS and a budget $n_s \in \mathbb{N}$, sensor placement aims at finding an optimal subset of sensors $V_s \subset V_c$ of cardinality $|V_s| = n_s$.*

Optimality hereby depends on the task at hand and usually is related to the idea of deriving the behavior of the unobserved nodes $V_c \backslash V_s$ from the observable heads $\mathbf{h}_{V_s}$ at the sensor nodes. A common approach is to find the sensor nodes $V_s \subset V_c$ that optimize subsequent detection algorithms such as reducing the amount of undetected leakages in a WDS [18]. However, a disadvantage of such supervised modeling is the necessity of the labels, which might be difficult to observe, and specificity to the task.

Baselines. As a remedy, [9] propose a sensor placement methodology that solely relies on heads $\mathbf{h}$ based on different demands $\mathbf{d}$. They use hand-crafted and discretized derivatives to estimate the sensitivity of a node $v \in V$, measured by its change in head h_v depending on the change in roughness and pipe burst, and integrate it into an OP. They solve the latter by applying the popular multicriterial genetic algorithm NSGA-II [3,6,7]. In addition, we implement two other baselines: For the first, we choose sensors randomly. The second is motivated by the spatial structure: We apply spectral graph clustering [8] based on the WDS structure and pipe attributes and choose one sensor node per cluster randomly.

3.3 Network Rehabilitation

In general, network rehabilitation aims at modifying an already existing WDS as a consequence of changing demands in the future. In close cooperation with water-domain experts, we define the task of adapting diameters to changing demands. The expected new demands cannot be satisfied by the WDS due to

too small pipe diameters. Therefore, some pipes need to be replaced by pipes with larger diameters while minimizing the changes and thus the required costs. Moreover, new diameters lead to new pressure heads that need to satisfy lower and upper pressure bounds for operational reasons.

Definition 3 (Diameter rehabilitation). *Given a set of consumer nodes $V_c \subset V$ in a WDS, expected new demands $\overline{\mathbf{d}}_i$ for $i = 1, ..., n$, which are higher than previous ones, current pipe diameters $\overline{\boldsymbol{\delta}} = (\overline{\delta}_e)_{e \in E} \in \mathbb{R}_+^{n_e}$, and lower and upper pressure bounds $\mathbf{h}_{V_c}^-, \mathbf{h}_{V_c}^+ \in \mathbb{R}_+^{n_c}$ on consumer nodes; diameter rehabilitation aims at finding new pipe diameters $\boldsymbol{\delta} = (\delta_e)_{e \in E} \in \mathbb{R}_+^{n_e}$ that obey the new demands and cause new pressure heads $\overline{\mathbf{h}}_i$ that obey the pressure bounds, i.e., $h_v^- \leq \overline{h}_{iv} \leq h_v^+$ for all $i = 1, ..., n$ and $v \in V_c$.*

Baselines. Currently, stochastic optimization algorithms are used to solve network rehabilitation tasks [15]. We will compare our approach with the popular multi-objective optimization scheme NSGA-II [3,6,7] mentioned before.

4 Background: PI-GCN Surrogate Model

Our work builds upon [2] that introduces a PI-GCN that is able to estimate physically correct states $(\tilde{\mathbf{h}}, \tilde{\mathbf{q}}, \tilde{\mathbf{d}})$ of a WDS given its heads $\mathbf{h}_{V_\mathrm{r}}$ at reservoir nodes and its demands $\mathbf{d} = \mathbf{d}_{V_\mathrm{c}}$ at consumer nodes (cf. *State Estimation* in section 3). It consists of a trainable part and an integrative part which corrects predicted values to obey hydraulic principles. It is trained based on the objective function $\mathcal{L}_{\text{PI-GCN}} = \mathcal{L}(\mathbf{d}, \hat{\mathbf{d}}) + \rho\mathcal{L}(\mathbf{d}, \tilde{\mathbf{d}}) + \delta\mathcal{L}(\hat{\mathbf{q}}, \tilde{\mathbf{q}})$, where $(\hat{\mathbf{d}}, \hat{\mathbf{q}})$ and $(\tilde{\mathbf{d}}, \tilde{\mathbf{q}})$ are the intermediate outputs from the GCN and the final outputs of the PI-GCN model, respectively. A detailed description is given in appendix A.2. Depending on the task to be solved, the inputs to the model are given or will be estimated.

5 Methodology

The main concept of this work is to leverage physics-informed gradients accessible through a fully differentiable surrogate model. This allows solving the downstream tasks introduced in section 3 by using gradient information (or more general, Jacobians) with respect to different parameters of the WDS.

In this work, we use the state estimation surrogate from [2] introduced in section 4. More precisely, after training the PI-GCN for its optimal parameters $\Theta_{\text{opt.}}$ as in [2], we fix these parameters $\Theta_{\text{opt.}}$ and introduce a new objective $f : \mathbb{R}^{n_{\text{in}}} \to \mathbb{R}^{n_{\text{out}}}$, $\mathbf{x} \mapsto f(\mathbf{x})$, which formalizes the downstream task. Consecutively, we use the differentiability of the function f, which is induced by the differentiability of the PI-GCN, to optimize the task-dependent parameter $\mathbf{x} \in \mathbb{R}^{n_{\text{in}}}$. A summary of the task-dependent network parameter $\mathbf{x}$ and the function f per task is displayed in table 1. Additionally, the general concept of leveraging gradients independent of WDSs is displayed in figure 1.

Table 1: Choices for the parameter $\mathbf{x} \in \mathbb{R}^{n_{\text{in}}}$ and function $f : \mathbb{R}^{n_{\text{in}}} \to \mathbb{R}^{n_{\text{out}}}$.

Task	Task-dep. parameter	Objective function
Sparse-to-dense pressure estimation	$\mathbf{d}$	eq. (1)
Sensor placement	$\mathbf{d}$	$\tilde{\mathbf{h}} = \tilde{\mathbf{h}}(\mathbf{d})$
Network rehabilitation	$\boldsymbol{\delta}$	eq. (9)

5.1 Sparse to Dense Pressure Estimation

Given the assumptions from definition 1, we utilize the differentiability of a trained PI-GCN [2] to estimate the pressure heads $\mathbf{h}$ in the whole network.

Required data. To leverage this model, we assume the availability of *prior or realistic* consumer demands $\mathbf{d}_i$ for $i = 1, ..., n$. Since the reservoir heads are given by definition of the sparse-to-dense pressure estimation task, we do not require further data to leverage this model.

Objective function. For this task, we choose the input demands $\mathbf{d} \in \mathbb{R}_+^{n_c}$ as the task-dependent network parameter. In order to find optimized demands $\mathbf{d}$ that produce outputs $(\tilde{\mathbf{h}}, \tilde{\mathbf{q}}, \tilde{\mathbf{d}})$ such that the heads $\tilde{\mathbf{h}}$ suit the sparse heads $\mathbf{h}_{V_\text{r} \cup V_\text{s}}$ on reservoir and sensor nodes, we choose $f(\mathbf{x}) = f(\mathbf{d})$ to be a suitable loss function with respect to which the demands $\mathbf{d}$ are optimized:

$$f(\mathbf{x}) = f(\mathbf{d}) = \mathcal{L}_{\text{S2D}}(\mathbf{d}) = \sum_{v \in V_\text{r} \cup V_\text{s}} |\tilde{h}_v(\mathbf{d}) - h_v|^2. \tag{1}$$

Minimizing this loss through back-propagation yields WDS states $\mathbf{d} \mapsto (\tilde{\mathbf{h}}, \tilde{\mathbf{q}}, \tilde{\mathbf{d}})$ such that $\tilde{\mathbf{h}}_{V_\text{r} \cup V_\text{s}} \approx \mathbf{h}_{V_\text{r} \cup V_\text{s}}$ holds. The model's final output $\tilde{\mathbf{h}}$ is the solution to the sparse-to-dense pressure estimation task as defined in definition 1.

Remark 1 (Availability of realistic demands). Our methodology is of high significance for other domain-related tasks: It does not only yield $\tilde{\mathbf{h}}$ as the solution to the sparse-to-dense pressure estimation task, but the whole WDS states $(\tilde{\mathbf{h}}, \tilde{\mathbf{q}}, \tilde{\mathbf{d}})$. Especially, we also obtain *realistic* demands $\tilde{\mathbf{d}}$ and flows $\tilde{\mathbf{q}}$, which are required for other downstream tasks, such as sensor placement (subsection 5.2).

Initial parameters. In order to optimize the loss $\mathcal{L}_{\text{S2D}}(\mathbf{d})$ with respect to (w.r.t.) the demands $\mathbf{d}$ via back-propagation, we need to initialize a starting point $\mathbf{d}_0$ and choose a reservoir and sensor head observation $\mathbf{h}_{V_\text{r} \cup V_\text{s}}$ for a fixed set of sensors V_s. However, as the sparse-to-dense pressure estimation task is ill-posed, different initializations $\mathbf{d}_0$ can lead to different solutions $\tilde{\mathbf{h}}$ still satisfying $\tilde{\mathbf{h}}_{V_\text{r} \cup V_\text{s}} = \mathbf{h}_{V_\text{r} \cup V_\text{s}}$.

Therefore, in order to obtain statistically significant results, we will consider different statistics over multiple solutions $\tilde{\mathbf{h}}_{li}$ via Monte Carlo sampling. The multiple solutions are obtained by different initializations $\mathbf{d}_{li}$ based on multiple sensor readings $\mathbf{h}_{lV_\text{r} \cup V_\text{s}}$ for $l = 1, ..., n_0$ and a suitably chosen subset $\{\mathbf{d}_i \mid i \in I_l\}$ of the demands $\mathbf{d}_i$ for $i = 1, ..., n$ (with I_l a subset of $\{1, ..., n\}$ such that $|I_l| = n_1 < n$ holds). We give detailed descriptions in appendix B.1.

Evaluation. We evaluate the solutions $\tilde{\mathbf{h}}_{li}$ of the sparse-to-dense pressure estimation task for $l = 1, ..., n_0$ and $i \in I_l$ by measuring the mean relative absolute error (MRAE) between these solutions and the true heads $\mathbf{h}_{lV_\mathrm{r} \cup V_\mathrm{s}}$ over the reservoir and sensor nodes, i.e.,

$$\mathrm{MRAE}_{\mathrm{S2D1}} = \frac{1}{n_0 n_1 (n_\mathrm{r} + n_\mathrm{s})} \sum_{l=1}^{n_0} \sum_{i \in I_l} \sum_{v \in V_\mathrm{r} \cup V_\mathrm{s}} \frac{|\tilde{h}_{liv} - h_{lv}|}{h_{lv}}. \tag{2}$$

Additionally, we compare our solutions $\tilde{\mathbf{h}}_{li}$ over all nodes to the solutions $\hat{\mathbf{h}}_{li}$ of EPANET, representing the ground truth since real-world data is not available. We obtain the latter by inputting the demands $\tilde{\mathbf{d}}_{li}$ for $l = 1, ..., n_0$ and $i \in I_l$ that are also output of the model of [2] to EPANET. We record the MRAE between these two solutions:

$$\mathrm{MRAE}_{\mathrm{S2D2}} = \frac{1}{n_0 n_1 n_\mathrm{n}} \sum_{l=1}^{n_0} \sum_{i \in I_l} \sum_{v \in V} \frac{|\tilde{h}_{liv} - \hat{h}_{liv}|}{\hat{h}_{liv}}. \tag{3}$$

5.2 Sensor Placement

Given the assumptions from definition 2, we leverage information from the gradients of a trained PI-GCN of [2] together with information theory to find an optimal subset of sensors $V_\mathrm{s} \subset V_\mathrm{c}$.

Required data. To leverage this model, we assume the availability of *realistic* reservoir heads and consumer demands $\mathbf{h}_{iV_\mathrm{r}}$ and $\mathbf{d}_i$ for $i = 1, ..., n$, respectively.

Gradients. Intuitively, as sensors at nodes measure the heads, sensors are needed at places where heads are sensitive to a change in demands in the WDS. Therefore, for the sensor placement task, we again choose the task-dependent network parameter as the input demands $\mathbf{d} \in \mathbb{R}_+^{n_\mathrm{c}}$. Choosing the function $f(\mathbf{x}) = (f_v(\mathbf{d}))_{v \in V_\mathrm{c}} = \tilde{\mathbf{h}}_{V_\mathrm{c}} \in \mathbb{R}_+^{n_\mathrm{c}}$ as the heads on consumer nodes outputted by the PI-GCN, the sensor placement relates to the question for which node $v \in V_\mathrm{c}$, the function f_v changes most w.r.t. any change of some demands $\mathbf{d}$. This, in turn, translates to the question which function f_v has the largest gradient $\nabla_\mathbf{d} f_v(\mathbf{d}) = (\mathbf{J}_\mathbf{d} f(\mathbf{d}))_v$. over all consumer nodes $v \in V_\mathrm{c}$.

As PI-GCN is non-linear, the Jacobians $\mathbf{J}_\mathbf{d} f(\mathbf{d})$ might differ for different input demands $\mathbf{d}$, we therefore seek the subset of sensor nodes V_s that optimizes the mean gradient norm over samples $\mathbf{d}_i$ for $i = 1, ..., n$ and sensor nodes $v \in V_\mathrm{s}$:

$$\begin{cases} \underset{V_\mathrm{s} \subset V_\mathrm{c}, |V_\mathrm{s}| = n_\mathrm{s}}{\operatorname{argmax}} \dfrac{1}{n n_\mathrm{s}} \displaystyle\sum_{i=1}^{n} \sum_{v \in V_\mathrm{s}} ||\nabla_\mathbf{d} f_v(\mathbf{d}_i)||_2. \end{cases} \tag{4}$$

Limitations of plain gradients. OP (4) comes with two drawbacks: First, the optimization over subsets $V_\mathrm{s} \subset V_\mathrm{c}$ with budget n_s requires $\binom{n_\mathrm{c}}{n_\mathrm{s}}$ calls of the

loss functions, which in most real-world scenarios makes it computationally expensive. Second, the head distributions of neighboring nodes or nodes in close proximity are often highly correlated. Therefore, neighboring nodes $v, u \in V_\mathrm{c}$ usually have similar gradients $\nabla_\mathbf{d} f_v(\mathbf{d}_i)$ and $\nabla_\mathbf{d} f_u(\mathbf{d}_i)$ for any $i \in \{1, ..., n\}$. Due to their correlation, however, the information provided by the two nodes is redundant. In other words, the two nodes have high mutual information.

Mutual information and its approximation. We solve the second problem by simultaneously decreasing the mutual information between the random variables H_v and H_u that are distributed according to the pressure heads at corresponding nodes $v, u \in V_\mathrm{c}$. Since these heads are real-valued, the mutual information is formally defined by their continuous densities $\varphi_v, \varphi_u, \varphi_{vu}$ (definition 6 in appendix A.3). In practice, the densities $\varphi_v, \varphi_u, \varphi_{vu}$ are unknown. We solve this problem by approximating them by step functions. The approach is based on the idea of using normalized histograms with $n_\mathrm{b} \in \mathbb{N}$ bins $A_j := [a_j, a_{j+1})$ for $j = 0, ..., n_\mathrm{b} - 2$ and $A_{n_\mathrm{b}-1} = [a_{n_\mathrm{b}-1}, a_{n_\mathrm{b}}]$. The histograms are created based on the Monte Carlo samples $f(\mathbf{d}_i) = \tilde{\mathbf{h}}_{iV_\mathrm{c}} = (\tilde{h}_{iv})_{v \in V_\mathrm{c}}$ for $i = 1, ..., n$.

As the focus of this work is not on the approximation of mutual information, the definition and a detailed derivation of the approximation of the mutual information of two continuously distributed random variables based on the approximated densities are given in appendix A.3. The main result of appendix A.3 is the following theorem, which we will make use of in the experiments.

Theorem 1 (Approximated mutual information). *In the setting of definition 5 and 7 (definition of density approximation and mutual information, respectively),*

$$\hat{I}(\mathrm{H}_v, \mathrm{H}_u) = \sum_{j_1=0}^{n_b-1} \sum_{j_2=0}^{n_b-1} p_{j_1,j_2}(v,u) \cdot \log\Big(\frac{p_{j_1,j_2}(v,u)}{p_{j_1}(v) \cdot p_{j_2}(u)}\Big) \qquad \textit{with}$$

$$p_{j_1,j_2}(v,u) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{A_{j_1} \times A_{j_2}}(\tilde{h}_{iv}, \tilde{h}_{iu}),$$

$$p_j(v) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{A_j}(\tilde{h}_{iv}) \qquad \textit{holds.}$$

Theorem 1 states that the approximated mutual information $\hat{I}(\mathrm{H}_v, \mathrm{H}_u)$ of two real-valued random variables with unknown densities equals the mutual information of the discrete probability distribution given by the relative amount of sampled observations $\tilde{\mathbf{h}}_{iV_\mathrm{c}}$ for $i = 1, ..., n$ in the discretized bins A_j and $A_{j_1} \times A_{j_2}$ for $j, j_1, j_2 \in \{0, ..., n_\mathrm{b} - 1\}$. The graph of these discretized densities corresponds to nothing more but the normalized histograms obtained by the Monte Carlo samples $\tilde{\mathbf{h}}_{iV_\mathrm{c}}$. Visualizations of such approximated densities will be presented in subsection 6.2.

By theorem 3 in appendix A.3, the approximated mutual information converges towards the true mutual information when the number of observed samples are chosen as $n \geq {n_\mathrm{b}}^2$ and the number of bins n_b goes to infinity.

Final algorithm. A natural step is to extend OP (4) by the mean mutual information, i.e.,

$$\left\{ \underset{V_\mathrm{s} \subset V_\mathrm{c}, |V_\mathrm{s}|=n_\mathrm{s}}{\operatorname{argmax}} \frac{1}{n n_\mathrm{s}} \sum_{i=1}^{n} \sum_{v \in V_\mathrm{s}} ||\nabla_{\mathbf{d}} f_v(\mathbf{d}_i)||_2 - \frac{1}{n_\mathrm{s}} \sum_{v,u \in V_\mathrm{s}} \hat{I}(\mathrm{H}_v, \mathrm{H}_u) \right. \tag{5}$$

between each two sensor nodes $v, u \in V_\mathrm{s}$. However, testing this for each possible subset $V_\mathrm{s} \subset V_\mathrm{c}$ would also be computationally expensive. We solve this problem by using the algorithm of [18] in conjunction with our per-node objectives. More precisely, we define our objectives (cf. eq. (5)) per node $v \in V_\mathrm{c}$ as

$$\mathcal{L}_1(v) = \frac{1}{n} \sum_{i=1}^{n} ||\nabla_{\mathbf{d}} \tilde{h}_{iv}||_2, \quad \mathcal{L}_2(v) = \frac{1}{n_\mathrm{s}} \sum_{u \in V_\mathrm{s}} \hat{I}(\mathrm{H}_v, \mathrm{H}_u) \tag{6}$$

instead of considering objectives per subset $V_\mathrm{s} \subset V$. Consequently, given a hyperparameter $\lambda \in [0, 1]$, Algorithm 1 iteratively optimizes OPs consisting of these two normalized losses to iteratively add a node to the sensor node set V_s.

Algorithm 1 Sensor placement algorithm

In: Consumer nodes V_c, budget n_s, sampled heads $\tilde{\mathbf{h}}_{iV_\mathrm{c}}$, hyperparameter $\lambda \in [0, 1]$.

$V_\mathrm{s} = \emptyset$
$V_\mathrm{s} \longleftarrow \underset{v \in V_\mathrm{c}}{\operatorname{argmax}} \ \mathcal{L}_1(v)$
while $|V_\mathrm{s}| < n_\mathrm{s}$ **do**
  for $v \in V_\mathrm{c} \setminus V_\mathrm{s}$ **do**
    Compute $\mathcal{L}_1(v), \mathcal{L}_2(v)$
  end for
  $V_\mathrm{s} \longleftarrow \underset{v \in V_\mathrm{c} \setminus V_\mathrm{s}}{\operatorname{argmax}} \ \lambda \ \frac{\mathcal{L}_1(v) - \min_{\mathcal{L}_1}}{\max_{\mathcal{L}_1} - \min_{\mathcal{L}_1}} - \ (1 - \lambda) \ \frac{\mathcal{L}_2(v) - \min_{\mathcal{L}_2}}{\max_{\mathcal{L}_2} - \min_{\mathcal{L}_2}}$
end while

Out: Sensor nodes V_s.

In short, our method combines the idea of using derivatives as a measure of sensitivity per node and mutual information as a measure of redundancy between sensor nodes. Unlike other approaches, we have direct access to the derivatives and do not depend on any downstream task. The algorithm's output V_s is the solution to the sensor placement task as defined in definition 2.

Evaluation. The quality of sensor placement is usually evaluated based on different downstream tasks in the WDS, such as leakage detection [18]. However, as elaborated above, a strength of our approach is the independence on downstream tasks. This is relevant in cases where the labels of the corresponding downstream tasks are not reliable and might not correspond to the ground truth.

At the same time, without the availability of a downstream task, this makes the evaluation of sensor placement itself a non-trivial task. In general, we expect

the sensors to observe as much information as possible within the system, or equivalently, not observing redundant information. Therefore, we propose the mean mutual information over sensor nodes

$$\hat{I}_{\mathrm{SP}} = \frac{1}{n_{\mathrm{s}}(n_{\mathrm{s}}-1)} \sum_{v \in V_{\mathrm{s}}} \sum_{u \in V_{\mathrm{s}} \setminus \{v\}} \hat{I}(\mathrm{H}_v, \mathrm{H}_u) \tag{7}$$

as a downstream-task-independent evaluation measure (cf. theorem 1).

For a downstream-task-dependent evaluation measure, we observe that the task of sparse-to-dense pressure estimation is based on a fixed set of sensor nodes V_{s}. Therefore, we can use the solutions $\tilde{\mathbf{h}}_{li}$ from the sparse-to-dense pressure estimation task for $l = 1, ..., n_0$ and $i \in I_l$ (cf. subsection 5.1) to evaluate different sensor placements. For a fixed observation of reservoir and sensor heads $\mathbf{h}_{lV_{\mathrm{r}} \cup V_{\mathrm{s}}}$ for $l \in \{1, ..., n_0\}$ and for a fixed node $v \in V$, a better set of sensors V_{s} will lead to solutions $(\tilde{h}_{liv})_{i \in I_l}$ more similar to each other, or in other words, to a less surprising underlying distribution. In this case, we aim to minimize its entropy. We compute the entropy similar to how we approximate the mutual information in this subsection. For more details, we refer to appendix B.2. Up to a constant, the discrete entropy is given by

$$\hat{E}(\mathrm{H}_{lv}) = -\sum_{j=0}^{n_{\mathrm{b}}-1} p_j(v) \cdot \log\left(p_j(v)\right)$$

with $p_j(v)$ as defined in theorem 1 (or theorem 5 in appendix A.3), but using the new heads $\tilde{\mathbf{h}}_{li}$ for $i \in I_l \subset \{1, ..., n\}$ instead of the heads $\tilde{\mathbf{h}}_i$ for $i = 1, ..., n$.

Finally, to summarize the entropy, we report the mean $\hat{E}_{\mathrm{SP1}}$ and $\hat{E}_{\mathrm{SP2}}$ over all reservoir and sensor observations $l = 1, ..., n_0$ as well as reservoir and sensor nodes $V_{\mathrm{r}} \cup V_{\mathrm{s}}$, and all nodes V, respectively:

$$\hat{E}_{\mathrm{SP1}} = \frac{1}{n_0 n_{\mathrm{r}}} \sum_{l=1}^{n_0} \sum_{v \in V_{\mathrm{r}} \cup V_{\mathrm{s}}} \hat{E}(\mathrm{H}_{lv}), \quad \hat{E}_{\mathrm{SP2}} = \frac{1}{n_0 n_{\mathrm{n}}} \sum_{l=1}^{n_0} \sum_{v \in V} \hat{E}(\mathrm{H}_{lv}). \tag{8}$$

5.3 Network Rehabilitation

So far, we demonstrated how the gradients of a surrogate model can be used to optimize nodal parameters. In the context of network rehabilitation, we want to optimize diameters, i.e. edge parameters that we can optimize with our method as well. Given the assumptions from definition 3, we utilize the differentiability of a trained PI-GCN of [2] to find new pipe diameters $\boldsymbol{\delta}$.

Required data. To leverage this model, we assume the availability of *realistic* reservoir heads $\mathbf{h}_{iV_{\mathrm{r}}}$ for $i = 1, ..., n$. Since the demands are given per definition of the network rehabilitation task, we do not require further data.

Objective function. Using the future demands $\overline{\mathbf{d}}_i$ as inputs to the model, we obtain heads $\overline{\mathbf{h}}_i := \tilde{\mathbf{h}}_i$ for $i = 1, ..., n$ which should obey the given pressure constraints. The network rehabilitation task can then be formulated as the following OP:

$$\min_{\boldsymbol{\delta}=(\delta_e)_{e\in E}\in\mathbb{R}_+^{n_e}} \sum_{e\in E} |\delta_e - \overline{\delta}_e| \quad \text{s.t.} \quad \tilde{h}_{iv} \geq h_v^-,\ \tilde{h}_{iv} \leq h_v^+ \quad \forall v \in V_c, i = 1, ..., n.$$

This problem can be solved by the before-mentioned genetic algorithm NSGA-II [3,6,7], which will serve as a baseline.

A computationally less costly alternative can be obtained using our gradient-based concept: As the heads $\tilde{\mathbf{h}}_i$ are an output of the model, they depend on the WDS's diameters $\boldsymbol{\delta}$ which we want to optimize. Therefore, this time, we can choose the task-dependent network parameter to be these pipe diameters $\boldsymbol{\delta} = (\delta_e)_{e\in E} \in \mathbb{R}_+^{n_e}$. Consecutively, to make use of the differentiability of PI-GCN, we transform the OP into a single, (almost-everywhere) differentiable loss function which we optimize w.r.t. to the pipe diameters $\boldsymbol{\delta}$, where $\alpha, \beta, \gamma \in [0, 1]$ are hyperparameters:

$$\begin{aligned} f(\mathbf{x}) = f(\boldsymbol{\delta}) = \mathcal{L}_{\text{NR}}(\boldsymbol{\delta}) = \alpha \cdot \mathcal{L}_{\text{PI-GCN}}(\boldsymbol{\delta}) + \beta \cdot \sum_{e\in E} |\overline{\delta}_e - \delta_e| \\ - \gamma \cdot \frac{1}{n} \sum_{i=1}^{n} \sum_{v\in V} (\max\{h_v^+ - \tilde{h}_{iv}(\boldsymbol{\delta}), 0\} + \max\{\tilde{h}_{iv}(\boldsymbol{\delta}) - h_v^-, 0\}). \end{aligned} \tag{9}$$

Finally, after having trained the diameters $\boldsymbol{\delta}$, the optimized diameters correspond to the solution of the network rehabilitation task, delivering optimal changes required for each pipe in the WDS in order to satisfy future demands and pressure constraints as defined in definition 3.

Initial parameters. In order to optimize the loss $\mathcal{L}_{\text{NR}}(\boldsymbol{\delta})$ w.r.t. the diameter $\boldsymbol{\delta}$ via back-propagation, we need to initialize a starting point $\boldsymbol{\delta}_0$. In this case, we simply choose the old diameters $\boldsymbol{\delta}_0 = \overline{\boldsymbol{\delta}}$. They will not satisfy the pressure head constraints, leading to a large loss $\mathcal{L}_{\text{NR}}(\boldsymbol{\delta})$ that will be optimized iteratively.

Evaluation. We evaluate the solutions $\boldsymbol{\delta}$ of the network rehabilitation task by measuring the absolute cost (AC) and the mean absolute cost (MAC) between these solutions and the initial diameters $\overline{\boldsymbol{\delta}}$ over the edges, approximated by

$$\text{AC}_{\text{NR}} = \sum_{e\in E} |\overline{\delta}_e - \delta_e|, \quad \text{MAC}_{\text{NR}} = \frac{1}{n_e} \sum_{e\in E} |\overline{\delta}_e - \delta_e|. \tag{10}$$

6 Experiments

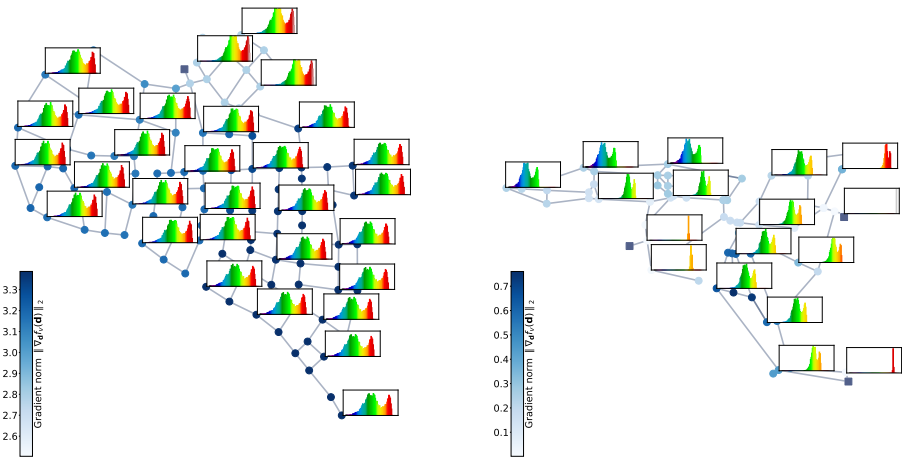
We conduct experiments on the same five WDS datasets as [2] do. Details of those can be found in appendix C.1. As we solve all tasks from section 3 utilizing trained models from [2], details on the training can be found in appendix C.2.

Table 2: $\text{MRAE}_{\text{S2D1}}$ (cf. eq. (2)) and $\text{MRAE}_{\text{S2D2}}$ (cf. eq. (3)) of the sparse-to-dense pressure estimation task on different WDSs.

Method	Hanoi	Fossolo	Pescara	Area-C	Zhi Jiang
	$\text{MRAE}_{\text{S2D1}}$				
mGCN	10.82 ± 2.33	0.62 ± 0.06	3.61 ± 0.83	1.82 ± 0.41	13.64 ± 0.49
ChebNet	1.75 ± 0.89	0.05 ± 0.03	0.28 ± 0.33	0.10 ± 0.09	0.21 ± 0.30
Ours	**0.10 ± 0.15**	**0.00 ± 0.00**	**0.16 ± 0.07**	**0.03 ± 0.04**	**0.17 ± 0.20**
	$\text{MRAE}_{\text{S2D2}}$				
mGCN	7.68 ± 1.97	0.42 ± 0.02	2.63 ± 0.88	1.63 ± 0.41	12.08 ± 1.26
ChebNet	1.38 ± 0.73	0.02 ± 0.01	0.66 ± 0.23	0.09 ± 0.08	0.25 ± 0.35
Ours	**0.14 ± 0.16**	**0.01 ± 0.00**	**0.23 ± 0.19**	**0.03 ± 0.04**	**0.17 ± 0.20**

6.1 Sparse to Dense Pressure Estimation

Per WDS, we sample $n_0 \cdot n_1 = 100 \cdot 48 = 480$ initialized demands $\mathbf{d}_0$ according to subsection 5.1. They correspond to $n_0 = 100$ different demands per $n_1 = 48$ different reservoir and sensor observations. Table 2 shows the $\text{MRAE}_{\text{S2D1}}$ (cf. eq. (2)) and $\text{MRAE}_{\text{S2D2}}$ (cf. eq. (3)) for different WDSs. The mGCN baseline model from [1] does not generalize well to unseen WDS attributes since it uses these pipe features as input. The ChebNet GCN model from [10] only uses sparse pressures as input and performs better than mGCN. We obtain significantly better results as compared to the baseline methods on all networks and both in comparison to the true sensor observations ($\text{MRAE}_{\text{S2D1}}$) and in comparison to the results of EPANET ($\text{MRAE}_{\text{S2D2}}$).



(a) Zhi Jang (b) Pescara

Fig. 2: Visualization of mean gradient norms in accordance with $\mathcal{L}_1$ and the approximated densities of head distributions that serve as a basis for $\mathcal{L}_2$.

6.2 Sensor Placement

We apply algorithm 1 to all five WDSs and for different configurations of the hyperparameter $\lambda \in [0, 1]$. A large λ puts more emphasis on $\mathcal{L}_1$, i.e. the gradients of the nodes, while a smaller λ puts more emphasis on $\mathcal{L}_2$, i.e., the mutual information between sensor nodes (cf. eq. (6)). To get an intuition for the two loss functions, figure 2 displays the mean gradient norms in accordance with $\mathcal{L}_1$ and the approximated densities of head distributions that serve as a basis for $\mathcal{L}_2$ (cf. paragraph *Mutual information and its approximation* in subsection 5.2) for the Zhi Jiang and Pescara WDSs.

Consequently, figure 3 displays different sensor configurations based on different such hyperparameters in the WDS Zhi Jiang. As described in subsection 5.2, $\lambda = 1$ that results in focusing on the gradients only leads to a cluster of sensors, located in an area where the largest gradients appear. In contrast, $\lambda = 0$ causes sensor placement according to their mutual information only (except for the first sensor, which is picked according to its gradient, cf. algorithm 1), which separates them all over the network. Typically, the mutual information between nodes close to a reservoir and nodes that are not close is low, hence some sensor nodes are placed close to a reservoir. Hyperparameters $\lambda \in (0, 1)$ lead to a solution that distributes the sensors according to their mutual information while considering the gradient sizes.

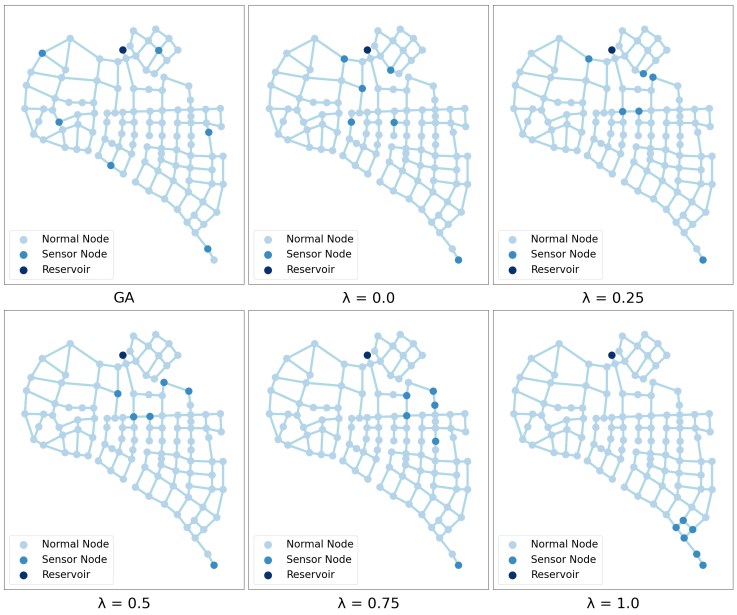


Fig. 3: Visualization of different sensor configurations based on different hyperparameters in the WDS Zhi Jiang.

Table 3: Mean mutual information $\hat{I}_{\mathrm{SP}}$ (cf. eq. (7)) for the sensor placement task for different sensor configurations in dependence of different hyperparameter λ and on different WDSs. Lowest values are highlighted in **bold** and second lowest in **gray**.

Method	Hanoi	Fossolo	Pescara	Area-C	Zhi Jiang
Random SP	**1.99 ± 0.00**	1.62 ± 0.00	1.64 ± 0.50	2.61 ± 0.10	2.76 ± 0.16
Clustering SP	2.59 ± 0.00	**1.09 ± 0.00**	1.73 ± 0.30	2.58 ± 0.10	2.74 ± 0.15
NSGA-II [9]	2.39 ± 0.00	1.97 ± 0.00	1.84 ± 0.24	2.56 ± 0.08	2.67 ± 0.10
Ours $\lambda = 0.00$	**0.00 ± 0.00**	**0.00 ± 0.00**	**0.24 ± 0.20**	**2.30 ± 0.12**	**2.61 ± 0.06**
Ours $\lambda = 0.125$	**0.00 ± 0.00**	**0.00 ± 0.00**	**0.24 ± 0.20**	**2.36 ± 0.12**	**2.61 ± 0.06**
Ours $\lambda = 0.25$	**0.00 ± 0.00**	**0.00 ± 0.00**	**0.24 ± 0.20**	**2.36 ± 0.12**	**2.64 ± 0.07**
Ours $\lambda = 0.50$	2.54 ± 0.00	1.97 ± 0.00	**0.85 ± 0.93**	2.65 ± 0.09	2.68 ± 0.07
Ours $\lambda = 0.75$	2.54 ± 0.00	1.97 ± 0.00	2.79 ± 0.17	2.67 ± 0.08	2.73 ± 0.13
Ours $\lambda = 1.00$	2.54 ± 0.00	2.14 ± 0.00	2.91 ± 0.08	3.24 ± 0.00	3.25 ± 0.02

Table 3 shows the downstream task-independent mean mutual information $\hat{I}_{\mathrm{SP}}$ (cf. eq. (7)) per WDS and hyperparameter λ. It can be seen that the sensor configurations with a smaller λ cause smaller mutual information, meaning that they observe more information from the network. For smaller λ, we also obtain better results as compared to the baseline methods.

Additionally, table 10 in appendix C.3 shows the downstream task-dependent mean entropy $\hat{E}_{\mathrm{SP1}}$ and $\hat{E}_{\mathrm{SP2}}$ (cf. eq. (8)) by WDS and per hyperparameter λ. Similar to the mutual information, the entropy $\hat{E}_{\mathrm{SP1}}$ on reservoir and sensor nodes is lower for smaller λ as compared to larger λ, and smaller than the baselines in four out of five cases. The method with the lowest entropy $\hat{E}_{\mathrm{SP2}}$ on all nodes differs among different WDSs and emphasizes that the evaluation of sensor placement on only one downstream task is not enough. Instead, future work should focus on the evaluation on several downstream tasks in order to investigate the average performance of sensor placement over all these tasks.

6.3 Network Rehabilitation

For each WDS, we consider multiple future demand collections and report the mean metrics as presented in subsection 5.3 over these collections. Table 4 shows the (mean) absolute cost $\mathrm{AC}_{\mathrm{NR}}$ and $\mathrm{MAC}_{\mathrm{NR}}$ (cf. eq. (10)) per WDS, which demonstrates the potential benefits of our approach as compared to the baseline.

Additionally, figure 4 visualizes the percentage change of diameters in response to a change in demands in the WDS Pescara. At nodes close to the increased demands, the pipe diameters get increased the most, whereas the diameters close to the leftmost reservoir remain mostly unchanged. From the perspective of construction cost this is generally beneficial, although actual cost calculations include many more parameters and considerations. Extending our loss function by such a term can be an interesting avenue for future research.

Table 4: (Mean) absolute error AC_{NR} and MAC_{NR} (cf. eq. 10) for the network rehabilitation task on different WDSs.

Method	Hanoi	Fossolo	Pescara	Area-C	Zhi Jiang
	AC_{NR}				
NSGA-II (NR)	365.6	801.8	1,596.0	233.6	121.3
Ours	**243.2**	**468.4**	**794.3**	**135.6**	**90.1**
	MAC_{NR}				
NSGA-II (NR)	10.8	13.8	16.1	2.1	0.7
Ours	**7.1**	**8.1**	**8.0**	**1.2**	**0.5**

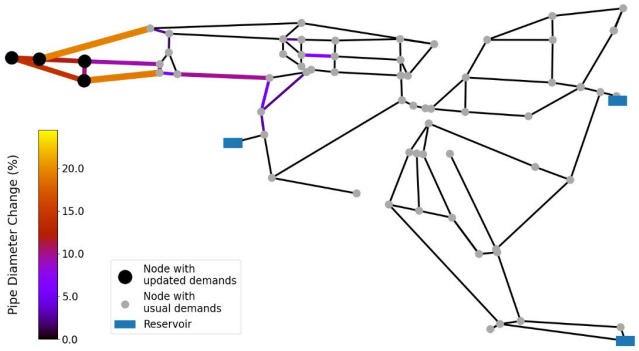


Fig. 4: Rehabilitation of the WDS Pescara. Black nodes experience a demand increase that leads to pressure dropping below the minimum pressure constraint (h_v^-). Our rehabilitation method increases the pipe diameters (colored lines) to allow for the additional flow required to satisfy h_v^-.

7 Conclusion and Future Work

In this work, we use the latest improvement in state estimation in WDSs using PI-GCN surrogate models as a starting point to solve several water-related downstream tasks such as sparse-to-dense pressure estimation, sensor placement, and network rehabilitation. This list is non-exhaustive and only demonstrates the possibilities that come with the differentiable state estimator presented by [2]. Moreover, this methodology extends to other differentiable surrogate models and is flexible w.r.t. available state information, allowing the backward pass of the surrogate model to be used as a digital twin. For example, sparse-to-dense pressure estimation requires the knowledge of prior consumer demands, which experts can usually provide. Meanwhile, sensor placement requires the knowledge of realistic reservoir heads and consumer demands, which can be obtained by the former method (cf. remark 1). Lastly, network rehabilitation requires only knowledge of realistic reservoir heads, which is reasonable to assume for

most situations. This allows immediate application of our solutions in practice, emphasizing the significance of our work to the water domain.

Limitations of this work are tightly coupled to the limitations of the PI-GCNs used for optimization. Scaling and the ability of our method to generalize are upper-bounded by the scaling of the surrogate model. However, our methods work with any differentiable physics-informed model and will benefit from further development of such. Moreover, all of our proposed methods are adaptable and can be improved given more information and data from the domain experts. For example, evaluating our sensor placement method on real-world data and different downstream tasks is an interesting avenue for future research. Incorporating an elaborate cost function into the task of network rehabilitation can lead to more structured results, as required by water experts.

Acknowledgments. We gratefully acknowledge funding from the European Research Council (ERC) under the ERC Synergy Grant Water-Futures (Grant agreement No. 951424).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ashraf, I., Hermes, L., Artelt, A., Hammer, B.: Spatial graph convolution neural networks for water distribution systems. In: Advances in Intelligent Data Analysis XXI. Springer Nature Switzerland (2023)
2. Ashraf, I., Strotherm, J., Hermes, L., Hammer, B.: Physics-informed graph neural networks for water distribution systems. Proceedings of the AAAI Conference on Artificial Intelligence **38** (2024)
3. Babayan, A.V., Savic, D.A., Walters, G.A.: Multi-objective optimization of water distribution system design under uncertain demand and pipe roughness. In: Topics on System Analysis and Integrated Water Resources Management. Elsevier (2007)
4. Brucherseifer, E., Winter, H., Mentges, A., Mühlhäuser, M., Hellmann, M.: Digital twin conceptual framework for improving critical infrastructure resilience. Automatisierungstechnik **69** (2021)
5. Candelieri, A., Ponti, A., Giordani, I., Archetti, F.: Lost in optimization of water distribution systems: Better call bayes. Water **14** (2022)
6. Creaco, E., Franchini, M., Walski, T.M.: Taking account of uncertainty in demand growth when phasing the construction of a water distribution network. Journal of Water Resources Planning and Management **141** (2015)
7. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: Nsga-ii. IEEE Transactions on Evolutionary Computation **6** (2002)
8. Donath, W., Hoffman, A.: Algorithms for partitioning graphs and computer logic based on eigenvectors of connection matrices. IBM Technical Disclosure Bulletin **15**(3), 938–944 (1972)
9. Ferreira, B., Antunes, A., Carriço, N., Covas, D.: Multi-objective optimization of pressure sensor location for burst detection and network calibration. Computers & Chemical Engineering **162** (2022)

10. Hajgató, G., Gyires-Tóth, B., Paál, G.: Reconstructing nodal pressures in water distribution systems with graph neural networks. arXiv preprint 2104.13619 (2021)
11. Hu, C., Li, M., Zeng, D., Guo, S.: A survey on sensor placement for contamination detection in water distribution systems. Wirel. Netw. **24** (2018)
12. Kerimov, B., Taormina, R., Tscheikner-Gratl, F.: Towards transferable metamodels for water distribution systems with edge-based graph neural networks. Water Research (2024)
13. Kreuzer, T., Papapetrou, P., Zdravkovic, J.: Artificial intelligence in digital twins—a systematic literature review. Data & Knowledge Engineering **151** (2024)
14. Magini, R., Moretti, M., Boniforti, M.A., Guercio, R.: A machine-learning approach for monitoring water distribution networks (wdns). Sustainability **15** (2023)
15. Mala-Jetmarova, H., Sultanova, N., Savic, D.: Lost in optimisation of water distribution systems? a literature review of system design. Water **10** (2018)
16. Nakka, R., Harursampath, D., Ponnusami, S.A.: A generalised deep learning-based surrogate model for homogenisation utilising material property encoding and physics-based bounds. Scientific Reports **13** (2023)
17. Rossman, L., Woo, H., Tryby, M., Shang, F., Janke, R., Haxton, T.: Epanet 2.2 user's manual, water infrastructure division. CESER (2020)
18. Santos-Ruiz, I., López-Estrada, F.R., Puig, V., Valencia-Palomo, G., Hernández, H.R.: Pressure sensor placement for leak localization in water distribution networks using information theory. Sensors **22** (2022)
19. Scott, D.W.: Multivariate Density Estimation: Theory, Practice, and Visualization. John Wiley & Sons (2015)
20. Truong, H., Tello, A., Lazovik, A., Degeler, V.: Graph neural networks for pressure estimation in water distribution systems. Water Resources Research **60** (2024)
21. Tsiami, L., Makropoulos, C., Savic, D.: Staged design of water distribution networks: A reinforcement learning approach. Engineering Proceedings **69** (2024)
22. Vrachimis, S., Kyriakou, M., Eliades, D., Polycarpou, M.: Leakdb: A benchmark dataset for leakage diagnosis in water distribution networks description of benchmark. In: Vol 1 of Proc., WDSA/CCWI Joint Conf. (2018)
23. Xing, L., Sela, L.: Graph neural networks for state estimation in water distribution systems: Application of supervised and semisupervised learning. Journal of Water Resources Planning and Management **148** (2022)