Sensor-Based Estimation of Contaminant Flow in Water Networks via Entropic Optimal Transport

Michele Mascherpa, Victor Molnö, Carsten Skovmose Kallesøe and Johan Karlsson

***Abstract*— We address the problem of estimating the spread of contaminants in a water network given sensors measurements. Modeling water flow as a Markov chain, we define states based on the volume of water in specific network sections (e.g., pipes). The goal is to infer the most probable pollution flow, and its source, given expected water dynamics and partial sensor observations. This leads to a large-scale optimization problem, formulated as a Schrödinger bridge with partial information, connected to entropy-regularized multimarginal optimal transport. The methodology is tested with data collected at the Smart Water Infrastructure Laboratory at the University of Aalborg.**

I. Introduction

The World Health Organization recognizes access to safe drinking-water as essential to health, a basic human right and a component of effective policy for health protection, see [12]. At the same time, failure to ensure drinking-water safety may expose the community to the risk of outbreaks of infectious diseases. The possible sources of contamination include pathogens and toxic chemicals, which can potentially enter water distribution systems as a result of malicious attacks, infrastructures aging or natural disasters, see e.g. [4]. The traditional laboratory-based methods for detection of pollution in water networks do not provide real-time public health protection. A significant measure that can be applied in order to mitigate the consequences of a contamination is the strategic placement of online water quality sensors, see [1]. Furthermore, recently, a large number of low cost sensors are becoming available, see [5].

Online sensing and real-time computations open up for development of new solutions for detection and estimation of pollution in water distribution systems, and with the aim of using the data produced by these sensors we propose a novel approach to estimate the spread of pollution in water networks, given the measurements from a set of sensors. Our approach is based on Markov modelling of the water diffusion in networks, where the states of the Markov chain are pipe segments. Once we model the water network as a Markov chain, we aim to find the most likely flow of pollution given the expected water flow and the sensors observations. This results in a large-scale optimization problem, which we formulate as a Schrödinger bridge problem with partial information. The Schrödinger bridge problem is a classical problem in statistical mechanics, see [6]. Given two observations of a particle distribution at two time instances, and a prior on the particle evolutions (in the classical setting this is the Brownian motion), the Schrödinger bridge describes the most likely evolution of the particle distributions. It turns out that the most likely evolution can be characterized as the one that minimizes the relative entropy to the prior, while also satisfying the observations. A discretized version of the Schrödinger bridge problem is based on modelling the particle evolutions as a Markov chain, see [3], [8]. The problem can be efficiently solved by an iterative scheme, which is linked to the so called Sinkhorn iterations for the optimal transport problem, see [2].

In this extended abstract, building on the work presented in [7], we utilize the discrete Schrödinger bridge to model the diffusion of pollution particles in a water network, in scenarios where only partial observations of the particle distributions are available. This reflects practical situations where pollution can only be observed at a limited number of sensor locations within the network. Furthermore, we generalize the approach by addressing the problem of identifying the pollution source in addition to modeling its spread. The proposed methodology is validated using data collected at the Smart Water Infrastructures Laboratory (SWIL) [11], a modular test facility at Aalborg University, Denmark, designed to emulate water distribution networks.

II. Problem formulation and main results

Consider a water network divided into n states, each representing a pipe-segment. We model the evolution of the pollution over $\mathcal{T}$ time instances, and we assume that each pollution particle evolves independently of each other and according to a Markov chain over these n states. We indicate with A^t the state transition matrix at time t, which is assumed to be known, and may be estimated from the water-flow. Moreover, let $M^t = [m^t]_{ij}$ describe the flow of pollution, i.e., m^t_{ij} indicates the amount of pollution that transitions from state i to state j between the times t and $t+1$. We assume that sensors at a few given locations (states) in the water network measure the amount of pollution, and these states are denoted by $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$. The measured pollution levels at these k states are described by vectors $\rho_t \in \mathbb{R}^k$, for $t \geq 1$. We introduce the matrix $B = [e_{\pi_1}, \ldots, e_{\pi_k}]^T \in \{0,1\}^{k\times n}$, where e_i is the i-th (column) vector of the canonical basis. Then, the matrix B has a nonzero column i if state i is known for all times. Let p and q be two nonnegative vectors or matrices of the same dimension. The normalized Kullback-Leibler (KL) divergence between p

This work is supported by KTH Digital Futures.

M. Mascherpa and J. Karlsson are with the Division of Optimization and Systems Theory, Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden. micmas@kth.se, johan.karlsson@math.kth.se. V. Molnö is with the Division of Decision and Control Systems, EECS, KTH Royal Institute of Technology, vmolno@kth.se. C.S. Kallesøe is with the University of Aalborg, Denmark and with Grundfos, Denmark, csk@es.aau.dk.

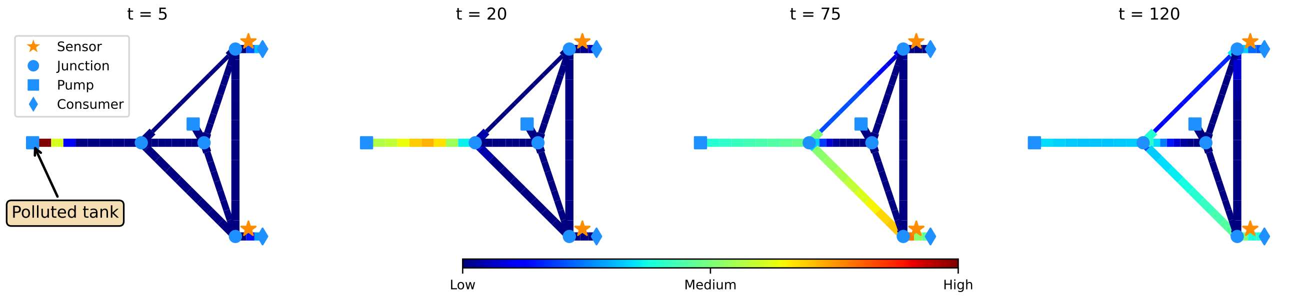

Fig. 1: Concentration of salt over time (seconds) in the emulated water network at SWIL.

from q is defined as $H(p|q) = \sum_i \left(p_i \log\left(\frac{p_i}{q_i}\right) - p_i + q_i \right)$, where $0 \log 0$ is defined to be 0.

We now formulate the problem of minimizing the entropy over time of the pollution flow M^t, with respect to the prior A^t, while respecting sensors measurements and conservation of mass at each time point.

$$\min_{M^{[0:\mathcal{T}-1]}} \sum_{t=0}^{\mathcal{T}-1} H(M^t \mid \mathrm{diag}(M^t \mathbb{1}) A^t) \tag{1a}$$

$$\text{s.t.} \quad BM^t \mathbb{1} = \rho_t \quad \text{for} \quad t = 0, \dots, \mathcal{T}-1, \tag{1b}$$

$$B(M^{\mathcal{T}-1})^T \mathbb{1} = \rho_{\mathcal{T}}, \tag{1c}$$

$$M^t \mathbb{1} = (M^{t-1})^T \mathbb{1} \quad \text{for} \quad t = 1, \dots, \mathcal{T}-1. \tag{1d}$$

An advantage of the Schrödinger bridge framework is that it requires only a few model assumptions, and its optimal solution is robust to errors in the assumed underlying model defined by the transition probabilities A^t, see [3], [9]. Problem (1) can be solved with an entropic proximal method [10]. The method adds an entropic penalty to the unconstrained part of the first marginal, relative to an initial estimate. This results in a problem with a fixed first marginal, whose dual can be solved using a Sinkhorn-type algorithm (see [7]). The solution is then used to update the estimate of the unconstrained entries, and the procedure is repeated until convergence.

III. Experiments

We consider the problem of estimating the source and the spread of pollution over a water network given the measurements from a set of sensors. We use data collected at SWIL, a modular test facility at the University of Aalborg. Pictures from the laboratory are presented in Fig. 2.

The setup contains two pumping stations, two consumer units, and seven 20m-pipes of different diameters. Pipes are also equipped with flow sensors, used to estimate the prior A^t. Salt is introduced in one pumping station, which is subsequently fed with clean water. Data from two conductivity sensors placed near the consumers are used. The network is discretized to a resolution of 2.5 meters. The optimal solution of problem (1) is presented in Fig.1 for different time points. Sensor measurements are matched up to a tolerance of 10^{-6}. We can observe how the method correctly identifies the contaminated pumping station as the source of pollution, also giving an estimate of its spread in time across the whole network.

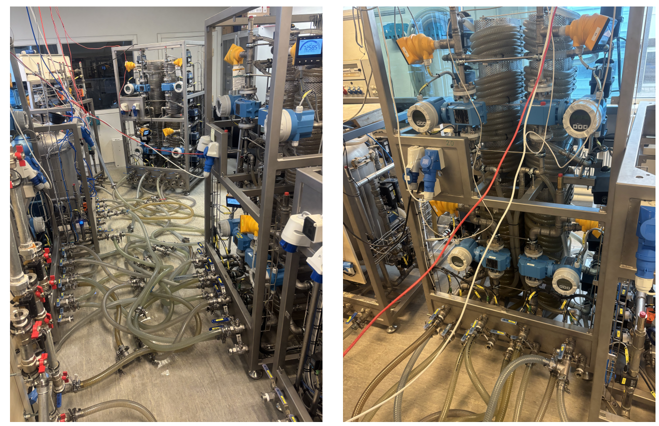

Fig. 2: Experimental setup at SWIL. Conductivity sensors (in yellow) are used to detect salinity.

References

[1] E. Z. Berglund, J. E. Pesantez, A. Rasekh, M. E. Shafiee, L. Sela, and T. Haxton. Review of modeling methodologies for managing water distribution security. *Journal of water resources planning and management*, 146(8):03120001, 2020.

[2] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26:2292–2300, 2013.

[3] I. Haasler, A. Ringh, Y. Chen, and J. Karlsson. Estimating ensemble flows on a hidden Markov chain. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 1331–1338. IEEE, 2019.

[4] R. Janke, M. E. Tryby, and R. M. Clark. Protecting water supply critical infrastructure: An overview. *Securing water and wastewater systems*, pages 29–85, 2014.

[5] T. P. Lambrou, C. C. Anastasiou, C. G. Panayiotou, and M. M. Polycarpou. A low-cost sensor network for real-time monitoring and contamination detection in drinking water distribution systems. *IEEE sensors journal*, 14(8):2765–2772, 2014.

[6] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *arXiv preprint arXiv:1308.0215*, 2013.

[7] M. Mascherpa, I. Haasler, B. Ahlgren, and J. Karlsson. Estimating pollution spread in water networks as a Schrödinger bridge problem with partial information. *European Journal of Control*, 2023.

[8] M. Pavon and F. Ticozzi. Discrete-time classical and quantum Markovian evolutions: Maximum entropy problems on path space. *Journal of Mathematical Physics*, 51(4):042104, 2010.

[9] M. Pavon, G. Trigila, and E. G. Tabak. The data-driven Schrödinger bridge. *Communications on Pure and Applied Mathematics*, 74(7):1545–1573, 2021.

[10] M. Teboulle. Entropic proximal mappings with applications to nonlinear programming. *Mathematics of Operations Research*, 17(3):670–690, 1992.

[11] J. Val Ledesma, R. Wisniewski, and C. S. Kallesøe. Smart water infrastructures laboratory: Reconfigurable test-beds for research in water infrastructures management. *Water*, 13(13):1875, 2021.

[12] WHO. Guidelines for drinking-water quality, 4th edition, incorporating the 1st addendum. 2017.