wetsus
european centre of excellence
for sustainable water technology

Machine learning approach for meta-genomic data interpretation in water process control

Asala Mahajna
asala.mahajna@wetsus.nl

Motivation

While water treatment processes in the Netherlands are of a very high standard, continuous threats from newly emerging micro-pollutants in (drinking-)water sources are still a challenge. Ensuring bio-stable water, which does not support the growth of microbes during water distribution is important for the prevention of aesthetic, operational and hygienic problems. The emergence of new genetic sequencing techniques that provide detailed information on the composition and response of the microorganisms, has inspired research to gain crucial in-situ information related to the microbial biogeography and occupancy-abundance dynamics in drinking water (illustrated in Figure 1, below). While meta-genomic data has been effectively collected, the analysis has been limited to correlational analysis with the underlying process data. However, an integrated systems biology approach which enables a predictive and adaptive modelling of the microbial dynamics in drinking water is still lacking [1,4,5] .

Technological challenge

The technological challenge is to choose and put to use a suitable machine learning approach that will automate the kinetic network reconstruction of carbon degradation and central metabolic pathways in water using metagenomics data. The resulting kinetic model will serve as a digital twin that will provide real-time state estimation of the microbial activities, as a proxy to drinking water quality sensors, and will enable operators to take timely actions to ensure a constant high quality of the treated water[2].

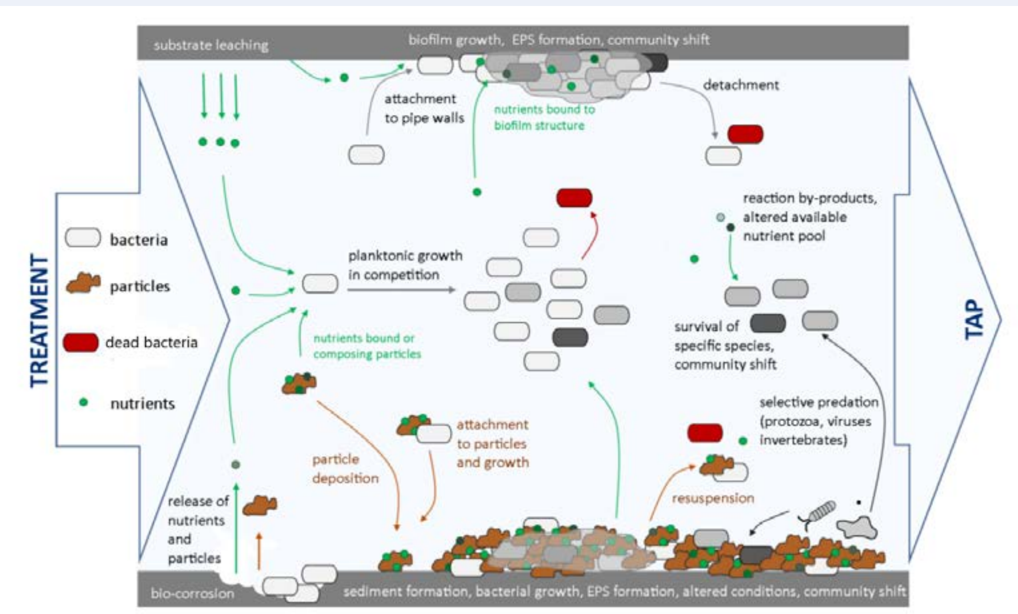

Fig 1. Overview of microbial dynamics in a water distribution pipe. Stimuli include hydraulic conditions, residence time, pipe material, and trophic chain.(modified from [3])

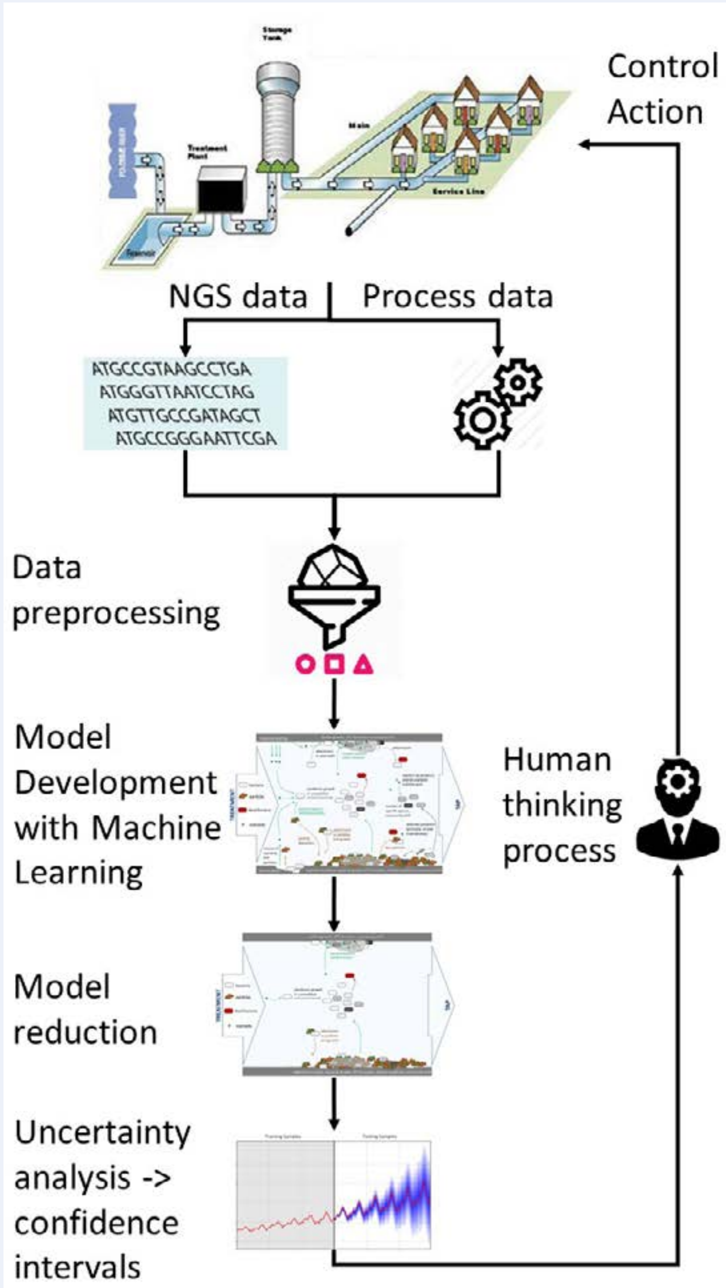

Fig 2. graphical abstract of the project

Research goals

1. Use machine learning tools to analyse meta-genomic and process data for automated kinetic network reconstruction of carbon degradation and central metabolic pathways in the water processing chain.
2. Development of a glass-box model that focuses on these metabolisms as proxies for microbial activities that affect water quality by applying kinetic model reduction techniques.
3. Development of model-based state estimators for the real-time prediction of water quality which includes the estimation of confidence intervals through uncertainty analysis.
4. Translate the model outcome to actions that can be taken to improve the water quality.

[1] Pinto, A., et al (2014). Spatial-Temporal Survey and Occupancy-Abundance Modeling To. Mbio,
[2] Negara P., M. A., et al B. “Next Generation Sequencing Analysis of Wastewater Treatment Plant Process Via Support Vector Regression,” 2019.
[3] Prest, E. et al (2016). Biological Stability of Drinking Water: Controlling Factors, Methods, and Challenges. Frontiers in Microbiology.
[4] Prest E, (2016) Long-Term Bacterial Dynamics in a Full-Scale Drinking Water Distribution System.
[5] Roeselers et al (2015). Microbial biogeography of drinking water: Patterns in phylogenetic diversity across space and time. Environmental microbiology.

Asala Mahajna M.Sc., dr. Inez J.T. Dinkla, prof.dr. Gert-Jan W. Euverink, prof.dr.ir. Karel J. Keesman, prof.dr.ir. Bayu Jayawardhana

WAGENINGEN UNIVERSITY & RESEARCH
university of groningen