Wastewater Disinfection: Modeling, Control and Risk Assessment

Jacopo Foschi

Politecnico di Milano

PhD in Infrastructure and Environmental Engineering

Supervisor:

prof. Manuela Antonelli

Co-supervisor

prof. Andrea Turolla

Tutor:

prof. Manuela Antonelli

XXXIV cycle

2018-2021

Abstract

In a global context of increasing human pressure on water resources and development of water reuse practices, disinfection of municipal wastewater has a primary role in controlling microbial contamination of natural water sources and water reuse systems. Operation of disinfection facilities are challenged by ever stricter limit at discharge on microbial indicators set by regulation and guidelines worldwide and by bottleneck of most chemical disinfectant, whose concentration in wastewater effluents cannot be raised, applying wide safety factors, without generating unacceptable toxic and ecotoxic impact in receiving waters. Ultraviolet (UV) disinfection gets around the use of chemicals, but still its energy consumption is not negligible when very high efficiency is needed and lamps end-of-life management is critical, since in most cases they are mercury-based. Such trade-off between health and environmental conflicting objectives demands for optimization and control of disinfectant, to achieve reliable compliance with very low target concentration of microbial indicators, while avoiding waste of chemicals and energy. This goal has been addressed on many aspects by old and recent research, which had to deal with intrinsic high variability of flow rate and quality of effluents to be disinfected and complexity of disinfection physical, chemical and microbiological mechanisms.

The aim of this PhD project is to conceive, calibrate and test models to optimize disinfection of wastewater. The disinfection process was studied combining lab-, pilot- and full-scale experiments and monitoring, to gather necessary data to develop and test predictive models of process disturbances and efficiency. Two main modeling approaches were explored. Data-driven machine learning models were used in cases where explicit description of physical world was not achievable or convenient, but data could be used to calibrate effective predictive models. Mechanistic models were preferred whenever a-priori knowledge of disinfection was sufficient to detailed modeling of involved hydraulic, chemical and microbiological phenomena. Models were deployed within control algorithms for real-time optimization of disinfectant dosage, compensating variability of effluent quality and flow rate.

Firstly, as essential precondition for real-time control of wastewater disinfection dosage, a “soft-sensor” approach is presented for virtual on-line monitoring of microbial indicator *Escherichia coli* in secondary and tertiary effluents at the inlet of disinfection units. Several linear and nonlinear regression models were studied, using conventional physical and chemical wastewater parameters as predictors of *E. coli* concentration. Results suggested that a neural network model is an effective

family of models for *E. coli* soft-sensing, catching complex nonlinear relationships between wastewater characteristics and microbial concentration. Potential benefits of the soft-sensor deployment were highlighted, simulating real-time optimization of chemical disinfectant dosage to compensate fluctuation in initial concentration of *E. coli* and estimating significant reductions of chemical consumption (about 66%) with respect to conventional off-line approaches for dosage design.

Secondly, a novel data-driven approach is illustrated for modeling and control of a full-scale UV disinfection of wastewater. The UV disinfection reactor of a large-scale wastewater treatment plant (WWTP), treating effluent for indirect reuse in agriculture, was monitored under various operating conditions and data were used to calibrate and test a black-box neural network model predicting *E. coli* disinfection efficiency given UV dose, flow rate and wastewater quality parameters affecting optical properties of water and disinfection mechanism. Results proved accuracy of the neural network model on test data and benefits of the combined use of the previously developed *E. coli* soft-sensor and the UV model for real time control of UV dose, which lead to an estimated saving of up to 66% of energy with respect to conventional off-line and experience-based dosage design.

Then, an original mechanistic model of peracetic acid (PAA) disinfection and its practical deployment and test at pilot-scale is presented. Lab-scale experiments were performed to calibrate sub-models of *E. coli* inactivation kinetics and PAA decay and how the latter is impacted by wastewater quality. Pilot-scale experiments were carried out to calibrate conceptual models of disinfection reactor hydrodynamics and validate disinfection model prediction accuracy and its effectiveness in controlling the process within a Model Predictive Control (MPC) algorithm. Validation of the PAA disinfection model confirmed the importance of considering dynamic nature of the process and impact of disturbances in prediction. Model deployment in MPC control at pilot scale proved effectiveness of the algorithm, with potential savings between 30% and 85% with respect to conventional *flow-paced* control.

Finally, reduction of health risk achieved by wastewater disinfection in an indirect agricultural reuse case study was estimated adopting the Quantitative Microbial Risk Assessment (QMRA) framework. Site-specific and literature data and models were integrated to describe fate and transport of reference pathogens of concern (*salmonella*, *norovirus*) from raw wastewater to moment of exposure by accidental ingestion by workers and during crop consumption. Results highlighted general low risk for the system under study and that a risk-based approach applied to

disinfection optimization allows to manage the process as part of a broader and complex system, including several pathogen inactivation barriers before human exposure, which must be completed by disinfection when and how much necessary. The study shew how QMRA can support decision-making in managing wastewater disinfection in a reuse system considering both health protection and environmental sustainability.

Keywords:
Wastewater disinfection; peracetic acid; UV disinfection; *Escherichia coli;* QMRA

Contents

Chapter 1: Introduction

1.1 Wastewater disinfection: state of art and current challenges

Wastewater, even after biological, physical and/or chemical treatments in wastewater treatment plants (WWTPs), contains relevant concentrations of pathogenic bacteria, viruses and protozoa of enteric origin, which can severely affect human health. Discharge of untreated or inadequately treated wastewater can indirectly impair sources of drinking water and bathing areas, potentially causing outbreaks of diseases like gastroenteritis and hepatitis. Waterborne outbreaks caused by fecal pollution are a major concern in many undeveloped and developing countries, where wastewater treatment facilities are inadequate or completely lacking. However, wastewater-originated fecal pollution leads to thousands of hospitalizations per year in developed countries too (Xagoraraki et al., 2014, Ma et al., 2022) .

Disinfection of municipal wastewater is designed to control pathogen concentration at the point of discharge of WWTPs, with the aim of minimizing concentration of enteric pathogens in receiving waters and protecting human health. While secondary and tertiary treatments usually lead to effluents containing 10^3 to 10^5 units/100 mL of a given pathogen (Inc. Metcalf & Eddy et al., 2013), disinfection can reduce this concentration by several order of magnitudes, potentially reaching almost null concentration (Mezzanotte et al., 2007, Antonelli et al., 2008, Rachmadi et al., 2020). Disinfection treatment is carried out with both chemical and physical processes. Chemical disinfection implies inactivation of pathogenic microorganisms by dosage of a chemical biocidal agent. Chlorine-based compounds, like chlorine gas (Cl_2), chloramines (NH_xCl_y) and sodium hypochlorite (NaOCl), are the oldest and best-established chemical disinfectants, with long proven high and wide spectrum biocidal effectiveness. Most common alternatives are ozone (O_3) (Morrison et al., 2022), mainly used as oxidants in wastewater treatment (Lim et al., 2022), and chlorine dioxide (ClO_2), even if it is not largely diffused (*inter alia*: Wang et al., 2007, Ayyildiz et al., 2009, Bischoff et al., 2012). Since all these disinfectants cause the formation of toxic and carcinogenic by-products and have high eco-toxicological impact, the use of organic peracids is spreading in more recent years. Among peracids, Peracetic Acid (CH_3CO_3H, PAA) and Performic Acid (CH_2O_3, PFA) are probably the main emerging chemical disinfectants. While they maintain similar disinfection efficiencies as conventional competitors, they lead to very low to null concentration of harmful by-products and have lower eco-toxicological impact. In case of PAA, DBPs are mostly carboxylic acids, while other potentially harmful by-products such as aldehydes, epoxides,

halogenated compounds and N-nitrosamines were detected at very low concentrations. (Domínguez Henao et al., 2018, Luukkonen & Pehkonen, 2017).

Physical disinfection systems are based on ultraviolet (UV) radiation or membranes and they have the general advantage of avoiding the use of chemicals and, thus, potential toxic and eco-toxic side effects. UV disinfection achieves pathogen inactivation by emission of radiation at wavelengths between 200 and 300 nm from immersed lamps. UV disinfection demonstrated as one of the most effective alternatives in virus inactivation. In more recent years, solar disinfection is developing and spreading, as cheaper solution, mainly for small-scale cases, exploiting natural UV radiation of sunlight (Bichai et al., 2012, Igoud et al., 2015). Membrane-based treatments with porous size lower than 1 µm, being micro-, ultra- and nanofiltration and reverse osmosis, can be considered as *de-facto* disinfection technologies, even if they do not achieve inactivation of pathogens but separate them from water. These solutions are expensive and critical to manage, usually designed to achieve multiple quality requirements in treated effluents, and rarely they are placed in treatment trains exclusively for pathogen reduction purposes (Ezugbe & Rathilal, 2020, Chen et al., 2021, Bera et al., 2022). For these reasons, they will be not further discussed in the present work.

Impact on inactivation efficiency of the combination of UV radiation and chemical disinfectants was widely explored, both as *ad-hoc* treatment for pathogen reduction and as positive side-effect of Advanced Oxidation Processes. In fact, combination of UV with hydrogen peroxide, ozone, titanium dioxide and PAA showed a synergistic effect, which increases the inactivation efficiency thanks to the production of high oxidation potential reactants, like hydroxyl radicals (OH•) (Chen et al., 2021). Electro-disinfection, sono-disinfection and natural-based solutions, like constructed wetlands, are other notable disinfection alternatives emerging in recent years (Collivignarelli et al., 2018).

In last decades, wastewater reuse is ever more spreading and developing as fundamental measure to compensate increasing water scarcity worldwide, driven by growing demographic pressure and climate change (Fito & Van Hulle, 2021, Ungureanu et al., 2020, Jeong et al., 2016). This is one of main current challenges for wastewater disinfection and gives it central importance in wastewater treatment trains. Direct, indirect and *de-facto* reuse of treated wastewater for potable, irrigation and industrial purposes potentially expose humans to pathogens by many routes, like ingestion, dermal contact and inhalation of aerosols, thus very high microbial quality standards must be achieved in treatment (Ofori et al., 2021, Kesari et al., 2021, Jaramillo & Restrepo, 2017). Latest guidelines and regulations worldwide are then setting ever more low compliance limits at the point of discharge on concentration of microbial indicators (NRMMC-EPHC-AHMC, 2006, The European Parliament

and the Council of the European Union, 2020, World Health Organization, 2006). The new Regulation (2020/741) of the European Union (The European Parliament and the Council of the European Union, 2020) on water reuse in agriculture is one of the most recent and with the most potential impact, setting new limit at discharge on the indicator microorganism *E. coli*, which are particularly strict in case of reuse for irrigation of crops eaten fresh and critical irrigation techniques like sprinkler irrigation (e.g. 10 or 100 CFU/100ml). Moreover, the Regulation explicitly promote a proactive management of water reuse systems based on risk assessment, which should lead to even stricter discharge limits in case characteristics of a specific reuse system imply higher potential exposure to pathogens.

Disinfection efficiency is driven by two main operating parameters, being the concentration of chemical disinfectant (or the UV radiation intensity) and contact time. While such high microbial concentration reduction is often practically achievable by sufficiently increasing chemical disinfectant concentration or UV radiation intensity, this is not always environmentally sustainable. High concentration of chemicals results in unfeasible toxicological and eco-toxicological impacts, preventing compliance of disinfection with both discharge of the effluent in surface waters and reuse (Domínguez Henao, Turolla, et al., 2018, Affek et al., 2021). In many cases, WWTP effluents eco-toxicity is monitored and regulated and practitioners face the challenge of optimizing disinfectant dosage to comply with both microbial and eco-toxicological limits. For example, in Italy, where case studies of this PhD work are located, limits on results from acute toxicity tests on Daphnia magna (or alternative indicator organisms like Selenastrum capricornutum) are defined by national law on wastewater discharge in surface waters and public sewer (D.LGS. 152/99, Attachment 5, Table 3. Last update in 2006). UV disinfection gets around the presence of chemicals, but the increase in number of UV lamps and radiation intensity could result in significant energy consumption, thus an environmental impact is still present. Moreover, most of UV lamps used in disinfection are mercury-based and their disposal is critical (Jones et al., 2018).

The presence of two conflicting health and environment-oriented objectives demands for optimization of the disinfection process, minimizing chemicals and energy to reliably achieve health protection target, while avoiding excessive environmental impact. This optimization perspective requires to consider wastewater disinfection as a process which dynamically varies with time and effluent quality and thus needs for control. These concepts are further discussed in paragraph 2.

The various disinfection alternatives previously summarized present both advantages and drawbacks and they fit differently to case studies according to the upstream treatment train, treatment goals, wastewater characteristics, downstream receiving waters type, and many more

factors. It has not been possible to identify one disinfection type dominating all alternatives, yet. Still, some options evidently get around the major drawbacks of formation of most harmful DBPs. As previously mentioned, this is the case of peracids, with PAA being the most widespread and best-established, and UV disinfection. For this reason, case studies on these two alternatives will be explored and discussed in the current work, in the attempt of providing modeling, optimization and control tools for more reliable and sustainable disinfection of wastewater.

1.2 Modeling and control of disinfection

As further discussed in following thematic chapters, wastewater disinfection efficiency, meant as the reduction in the pathogen or indicator microorganism of interest, depends on two main operating conditions. The first one is the "amount" of disinfectant microbes are exposed to, represented by disinfectant concentration (mg L^{-1}) or UV radiation intensity (mW cm^{-2}), in case of chemical and UV disinfection, respectively. The second one is contact time, meant as time microorganisms are exposed to, given disinfectant concentration or UV radiation.

While disinfectant dosage can be controlled, the process is affected by two main disturbances, which are wastewater flow rate and quality. Flow rate impacts on reactor hydrodynamics and determines contact time of each fluid particle, and it varies according to production of wastewater from municipalities and/or demand of treated effluent in case of reuse practices. At same time, several wastewater quality parameters impact on disinfection efficiency. First and straightforward, variability of the concentration of the microorganism of interest in the effluent to be disinfected determines variations in output concentration, at given disinfection operation (dosage and contact time). Then, concentrations of organics, suspended solids and metals can affect presence of disinfectant and how it varies in space and time. In fact, these wastewater quality parameters significantly enhance instantaneous oxidative demand and gradual decay of chemical disinfectants (*inter alia*: Haas & Karra, 1984, Domínguez Henao, et al., 2018) and affect how UV radiation spread out in the water media (*inter alia*: Carré et al., 2018). This topic is further discussed in chapters 4, 5 and 6.

Another main source of variability is inactivation kinetics, namely the relationship of disinfectant concentration (or UV radiation intensity) and contact time with reduction in microorganism concentration. In fact, it depends on type of microorganism, type of disinfectant, water matrix characteristics and the interactions among these three. Despite numerous studies propose many different inactivation models to describe inactivation kinetics in WWTP effluents in batch conditions (Gyürek & Frinch, 1998, Santoro et al., 2015, Domínguez Henao et al., 2018a),

most literature agrees on the common assumption that concentration of disinfectant (or UV radiation intensity) and contact time can be synthesized in a single operating parameter, the "dose", univocally determining microorganism inactivation, and propose disinfection models based on the dose concept (*inter alia*: Neumann et al., 2007, Santoro et al., 2015, Dominguez Henao et al., 2018, Ahmed et al., 2019). Overall, the higher is the dose, the higher is inactivation. First and most basic definition of the dose is the product of dosed disinfectant concentration (or emitted UV radiation intensity) and contact time. While this formulation is easy to compute and useful for preliminary design, it does not account how the presence of disinfectant can vary in space and time during the process. More recent and effective definition of the dose accounts for the loss of chemical disinfectant due to decay (Haas & Joffe, 1994) and, in case of UV disinfection, the reduction of actual UV radiation intensity due to absorption, reflection and refraction.

Several approaches were developed through the years in guidelines, regulations and scientific literature to model disinfection in continuous flow conditions, to predict reduction in target microorganisms and, accordingly, optimize and control disinfectant dosage. First, disinfection was modeled approximating it as a batch process, estimating reduction of the target microorganism by means of batch inactivation kinetics models relating dose to inactivation level. This approach was introduced and applied by the US Surface Water Treatment Rule in 1989 (later updated in 2006, USEPA, 2006). The Integrated Disinfection Design Framework (IDDF, Ducoste et al., 2001) marked a shift towards a site-specific approach, which included modeling of disinfectant decay, inactivation kinetics and reactor hydrodynamics of the case study of interest, leading to more accurate efficiency estimates and, consequently, guaranteeing better optimization of dosage. Still, IDDF models disinfection as a stationary and deterministic process. More recent scientific literature on disinfection modeling is focusing on describing the impact of disturbances variability and model uncertainty on the estimation of process efficiency (*inter alia*: Neumann et al., 2007, Santoro et al., 2015, Ahmed et al., 2019). Besides, research is pointing at modeling disinfection as a dynamic process, affected by time varying disturbances (*inter alia*: Manoli et al., 2019). More detailed discussion on the state of art of wastewater disinfection modeling and related research proposed in this work are reported in Chapter 3 to 6.

1.3 Estimation of wastewater disinfection impact: microbial risk assessment

As previously mentioned in paragraph 1, municipal wastewater disinfection primary goal is to protect human health avoiding spread of enteric pathogens in environment and water reuse systems. Monitoring pathogen concentration in raw and treated wastewater at point of discharge and

throughout wastewater reuse systems is a challenging task, requiring high expertise in microbiology, expensive technologies and time-consuming measuring techniques. Culture-based methods reliably measuring concentration of infective pathogen concentration exist only for a subset of pathogens of concern. In many cases, only surrogate species can be cultivated, while many species can be measured only by real-time Quantitative Polymerase Chain Reaction (qPCR). While recent development of qPCR allowed easier and broader monitoring of pathogens, still this method showed limitations in case of disinfection monitoring, overestimating actual infective concentration of pathogens after exposure to disinfectant (Girones et al., 2010, Rönnqvist et al., 2014). Monitoring efficiency and compliance of wastewater disinfection would need frequent and periodical measurement of all pathogens of concern, or a subset of reference pathogens, which is economically and practically unfeasible. Easy to measure and culturable indicator microorganisms are instead monitored as proxies of the presence of pathogens, and limits at the point of discharge for WWTP in guidelines and regulations are set as indicator microorganism concentrations (Toribio-Avedillo et al., 2021, McKee & Cruz, 2021). Examples are fecal coliforms and *Escherichia coli*, for bacteria, coliphages for viruses and *Giardia lamblia* and *Cryptosporidium parvum* for protozoa. While such indicators are useful for regulators and WWTP practitioners to optimize disinfection of wastewater, the actual beneficial impact of disinfection on human health should be verified estimating how different management of the process affects exposure to the actual source of health risk, being enteric pathogens. Probabilistic modeling of pathogen fate and transport from source to human exposure within a risk assessment framework can get around the inability of direct monitoring of pathogen species and link disinfectant dose control strategies to quantitative estimate of risk of disease by exposure to treated wastewater. Quantitative Microbial Risk Assessment (QMRA) (Haas et al., 2014) is the best-established and most widespread framework to fit this purpose. As further detailed in chapter 7, QMRA considers as reference a subset of pathogens of concern and use literature and site-specific data to model all phenomena involving inactivation, regrowth and/or transport and finally assess risk and burden of disease. Use of QMRA is of fundamental importance to assess health risk and support decision making in any wastewater reuse practice (Zhiteneva et al., 2020). Control of disinfectant dose can be optimized according to a risk-based approach, considering disinfection as part of a wider integrated reuse system. Moreover, estimating effective health risk resulting from given disinfection operating conditions allows to quantitatively assess trade-off between beneficial effects of disinfection and drawbacks coming from use of chemicals and energy consumption.

To summarize paragraphs 1.2 and 1.3, microbial indicators, like *E. coli*, are necessary to set discharge limits and monitoring disinfection compliance, and WWTP practitioners should take

advantage of accurate modeling of the process to reliably comply with target indicator concentration in the effluent. In parallel, risk assessment approaches like QMRA should be used to verify actual beneficial effect of disinfection, considering site-specific characteristics of wastewater reuse system under study, especially in case of indirect and *de-facto* reuse, where pathogen routes from discharge to exposure are long and complex. Disinfection treatment target should be then increased or reduced if health risk from reuse results as non-tolerable or if some chemicals or energy can be saved without concerning increase of health burden.

References

Affek, K., Muszyński, A., Doskocz, N., & Załęska-Radziwiłł, M. (2021). Ecotoxicological effects of disinfection of treated wastewate. Desalination and Water Treatment, 233, 190–198. https://doi.org/10.5004/dwt.2021.27549

Ahmed, Y. M., Ortiz, A. P., & Blatchley, E. R. (2019). Stochastic Evaluation of Disinfection Performance in Large-Scale Open-Channel UV Photoreactors. Journal of Environmental Engineering (United States), 145(10). https://doi.org/10.1061/(ASCE)EE.1943-7870.0001562

Antonelli, M., Mezzanotte, V., & Nurizzo, C. (2008). Wastewater disinfection by UV irradiation: Short and long-term efficiency. Environmental Engineering Science, 25(3), 363–373. https://doi.org/10.1089/ees.2006.0268

Ayyildiz, O., Ileri, B., Sanik, S., 2009. Impacts of water organic load on chlorine dioxide disinfection efficacy. J. Hazard. Mater. 168, 1092–1097. https://doi.org/10.1016/j.jhazmat.2009.02.153

Bera, S. P., Godhaniya, M., & Kothari, C. (2022). Emerging and advanced membrane technology for wastewater treatment: A review. Journal of Basic Microbiology, 62(3–4), 245–259. https://doi.org/10.1002/jobm.202100259

Bichai, F., Polo-López, M.I., Fernández Ibañez, P., 2012. Solar disinfection of wastewater to reduce contamination of lettuce crops by Escherichia coli in reclaimed water irrigation. Water Res. 46, 6040–6050. https://doi.org/10.1016/j.watres.2012.08.024

Bischoff, A., Cornel, P., Wagner, M., 2012. Choosing the most appropriate technique for wastewater disinfection - parallel investigation of four disinfection systems with different preceding treatment processes. Water Pract. Technol. 7. https://doi.org/10.2166/wpt.2012.054

Carré, E., Pérot, J., Jauzein, V., & Lopez-Ferber, M. (2018). Impact of suspended particles on UV disinfection of activated-sludge effluent with the aim of reclamation. Journal of Water Process Engineering, 22(January), 87–93. https://doi.org/10.1016/j.jwpe.2018.01.016

Chen, C., Guo, L., Yang, Y., Oguma, K., & Hou, L. an. (2021). Comparative effectiveness of membrane technologies and disinfection methods for virus elimination in water: A review. Science of the Total Environment, 801, 149678. https://doi.org/10.1016/j.scitotenv.2021.149678

Chen, Y. di, Duan, X., Zhou, X., Wang, R., Wang, S., Ren, N. qi, Ho, S.H., 2021. Advanced oxidation processes for water disinfection: Features, mechanisms and prospects. Chem. Eng. J. 409, 128207. https://doi.org/10.1016/j.cej.2020.128207

Collivignarelli, M. C., Abbà, A., Benigna, I., Sorlini, S., & Torretta, V. (2018). Overview of the main disinfection processes for wastewater and drinking water treatment plants. Sustainability (Switzerland), 10(1), 1–21. https://doi.org/10.3390/su10010086

Domínguez Henao, L., Cascio, M., Turolla, A., & Antonelli, M. (2018). Effect of suspended solids on peracetic acid decay and bacterial inactivation kinetics: Experimental assessment and definition of predictive models. Science of the Total Environment, 643, 936–945. https://doi.org/10.1016/j.scitotenv.2018.06.219

Domínguez Henao, L., Delli Compagni, R., Turolla, A., & Antonelli, M. (2018). Influence of inorganic and organic compounds on the decay of peracetic acid in wastewater disinfection. Chemical Engineering Journal, 337(October 2017), 133–142. https://doi.org/10.1016/j.cej.2017.12.074

Domínguez Henao, L., Turolla, A., & Antonelli, M. (2018). Disinfection by-products formation and ecotoxicological effects of effluents treated with peracetic acid: A review. Chemosphere, 213, 25–40. https://doi.org/10.1016/j.chemosphere.2018.09.005

Ducoste, J., Carlson, K., & Bellamy, W. (2001). The integrated disinfection design framework approach to reactor hydraulics characterization. Journal of Water Supply: Research and Technology - AQUA, 50(4), 245–261. https://doi.org/10.2166/aqua.2001.0021

Ezugbe, E. O., & Rathilal, S. (2020). Membrane technologies in wastewater treatment: A review. Membranes, 10(5). https://doi.org/10.3390/membranes10050089

Fito, J., & Van Hulle, S. W. H. (2021). Wastewater reclamation and reuse potentials in agriculture: towards environmental sustainability. Environment, Development and Sustainability, 23(3), 2949–2972. https://doi.org/10.1007/s10668-020-00732-y

Girones, R., Ferrús, M. A., Alonso, J. L., Rodriguez-Manzano, J., Calgua, B., de Abreu Corrêa, A., Hundesa, A., Carratala, A., & Bofill-Mas, S. (2010). Molecular detection of pathogens in water - The pros and cons of molecular techniques. Water Research, 44(15), 4325–4339. https://doi.org/10.1016/j.watres.2010.06.030

Haas, C. N., & Joffe, J. (1994). Decay. 28(7), 1367–1369.

Haas, C. N., & Karra, S. B. (1984). Kinetics of microbial inactivation by chlorine-II Kinetics in the presence of chlorine demand. Water Research, 18(11), 1451–1454. https://doi.org/10.1016/0043-1354(84)90016-2

Igoud, S., Souahi, F., Eddine Chitour, C., Amrouche, L., Moussaoui, A., Boumrar, H., Chekir, N., Chouikh, A., 2015. Efficiency evaluation of solar photolysis and solar photocatalysis processes used for the wastewater disinfection. Desalin. Water Treat. 53, 2049–2058. https://doi.org/10.1080/19443994.2013.862866

Inc. Metcalf & Eddy, Tchobanoglous, G., Stensel, D., Tsuchihashi, R., & Burton, F. (2013). Wastewater Engineering: Treatment and Resource Recovery (5th editio). McGraw-Hill.

Jaramillo, M. F., & Restrepo, I. (2017). Wastewater reuse in agriculture: A review about its limitations and benefits. Sustainability (Switzerland), 9(10). https://doi.org/10.3390/su9101734

Jeong, H., Kim, H., & Jang, T. (2016). Irrigation water quality standards for indirect wastewater reuse in agriculture: A contribution toward sustainablewastewater reuse in South korea. Water (Switzerland), 8(4). https://doi.org/10.3390/w8040169

Jones, C. H., Shilling, E. G., Linden, K. G., & Cook, S. M. (2018). Life Cycle Environmental Impacts of Disinfection Technologies Used in Small Drinking Water Systems. Environmental Science and Technology, 52(5), 2998–3007. https://doi.org/10.1021/acs.est.7b04448

Kesari, K. K., Soni, R., Jamal, Q. M. S., Tripathi, P., Lal, J. A., Jha, N. K., Siddiqui, M. H., Kumar, P., Tripathi, V., & Ruokolainen, J. (2021). Wastewater Treatment and Reuse: a Review of

its Applications and Health Implications. Water, Air, and Soil Pollution, 232(5). https://doi.org/10.1007/s11270-021-05154-8

Lim, S., Shi, J.L., von Gunten, U., McCurry, D.L., 2022. Ozonation of organic compounds in water and wastewater: A critical review. Water Res. 213, 118053. https://doi.org/10.1016/j.watres.2022.118053

Luukkonen, T., & Pehkonen, S. O. (2017). Peracids in water treatment: A critical review. Critical Reviews in Environmental Science and Technology, 47(1), 1–39. https://doi.org/10.1080/10643389.2016.1272343

Ma, J. Y., Li, M. Y., Qi, Z. Z., Fu, M., Sun, T. F., Elsheikha, H. M., & Cong, W. (2022). Waterborne protozoan outbreaks: An update on the global, regional, and national prevalence from 2017 to 2020 and sources of contamination. Science of the Total Environment, 806, 150562. https://doi.org/10.1016/j.scitotenv.2021.150562

Manoli, K., Sarathy, S., Maffettone, R., & Santoro, D. (2019). Detailed modeling and advanced control for chemical disinfection of secondary effluent wastewater by peracetic acid. Water Research, 153, 251–262. https://doi.org/10.1016/j.watres.2019.01.022

McKee, A. M., & Cruz, M. A. (2021). Microbial and Viral Indicators of Pathogens and Human Health Risks from Recreational Exposure to Waters Impaired by Fecal Contamination. Journal of Sustainable Water in the Built Environment, 7(2), 03121001. https://doi.org/10.1061/jswbay.0000936

Mezzanotte, V., Antonelli, M., Citterio, S., & Nurizzo, C. (2007). Wastewater Disinfection Alternatives: Chlorine, Ozone, Peracetic Acid, and UV Light. Water Environment Research, 79(12), 2373–2379. https://doi.org/10.2175/106143007x183763

Morrison, C.M., Hogard, S., Pearce, R., Gerrity, D., von Gunten, U., Wert, E.C., 2022. Ozone disinfection of waterborne pathogens and their surrogates: A critical review. Water Res. 214, 118206. https://doi.org/10.1016/j.watres.2022.118206

Neumann, M. B., von Gunten, U., & Gujer, W. (2007). Uncertainty in prediction of disinfection performance. Water Research, 41(11), 2371–2378. https://doi.org/10.1016/j.watres.2007.02.022

NRMMC-EPHC-AHMC. (2006). Overview of the Australian Guidelines for Water Recycling: Managing Health and Environmental Risks. Protection and Heritage Council, Australian Health, 389. http://www.ephc.gov.au/sites/default/files/WQ_AGWR_GL__Managing_Health_Environmental_Risks_Phase1_Final_200611.pdf

Ofori, S., Puškáčová, A., Růžičková, I., & Wanner, J. (2021). Treated wastewater reuse for irrigation: Pros and cons. Science of the Total Environment, 760. https://doi.org/10.1016/j.scitotenv.2020.144026

Rachmadi, A. T., Kitajima, M., Kato, T., Kato, H., Okabe, S., & Sano, D. (2020). Required Chlorination Doses to Fulfill the Credit Value for Disinfection of Enteric Viruses in Water: A Critical Review. Environmental Science and Technology, 54(4), 2068–2077. https://doi.org/10.1021/acs.est.9b01685

Rönnqvist, M., Mikkelä, A., Tuominen, P., Salo, S., & Maunula, L. (2014). Ultraviolet Light Inactivation of Murine Norovirus and Human Norovirus GII: PCR May Overestimate the Persistence of Noroviruses Even When Combined with Pre-PCR Treatments. Food and Environmental Virology, 6(1), 48–57. https://doi.org/10.1007/s12560-013-9128-y

Santoro, D., Crapulli, F., Raisee, M., Raspa, G., & Haas, C. N. (2015). Nondeterministic Computational Fluid Dynamics Modeling of Escherichia coli Inactivation by Peracetic Acid in Municipal Wastewater Contact Tanks. Environmental Science and Technology, 49(12), 7265–7275. https://doi.org/10.1021/es5059742

The European Parliament and the Council of the European Union. (2020). Regulation (EU) 2020/741 of 25 May 2020 on minimum requirements for water reuse. Official Journal of the European Union, 2019(April), L 177/32-L 177/55.

Toribio-Avedillo, D., Blanch, A. R., Muniesa, M., & Rodríguez-Rubio, L. (2021). Bacteriophages as fecal pollution indicators. Viruses, 13(6). https://doi.org/10.3390/v13061089

Ungureanu, N., Vlăduț, V., & Voicu, G. (2020). Water scarcity and wastewater reuse in crop irrigation. Sustainability (Switzerland), 12(21), 1–19. https://doi.org/10.3390/su12219055

USEPA. (2006). National Primary Drinking Water Regulations: Long term 2 Enhanced Surface Water Treatment Rule (40 CFR Parts 9, 141, and 142). Federal Register, 71(3), 654–786.

Wang, L., Hu, H., Wang, C., Koichi, F., 2007. Change in genotoxicity of wastewater during chlorine dioxide and chlorine disinfections and the influence of ammonia nitrogen. Front. Environ. Sci. Eng. China 1, 18–22. https://doi.org/10.1007/s11783-007-0003-7

World Health Organization. (2006). Safe Use of Wastewater , Excreta and Greywater Guidelines for the Safe Use of. World Health, II, 204. http://whqlibdoc.who.int/publications/2006/9241546832_eng.pdf

Xagoraraki, I., Yin, Z., & Svambayev, Z. (2014). Fate of Viruses in Water Systems. Journal of Environmental Engineering, 140(7), 04014020. https://doi.org/10.1061/(asce)ee.1943-7870.0000827

Zhiteneva, V., Hübner, U., Medema, G. J., & Drewes, J. E. (2020). Trends in conducting quantitative microbial risk assessments for water reuse systems: A review. Microbial Risk Analysis, 16(July), 100132. https://doi.org/10.1016/j.mran.2020.100132

Chapter 2: Design of the research

Ever more increasing water scarcity demands for protection of available water sources from microbial pollution originated by municipal wastewater. In this context, reuse of treated wastewater is developing as powerful solution to mitigate human water footprint, and in such practice control of microbial pollution is even more important and challenging. Disinfection of wastewater within WWTPs assumes then a central role and should be optimized to balance health protection with sustainable process operation, addressed by use of chemicals and energy consumption.

The objective of the PhD thesis is the development of models to support optimization of wastewater disinfection to achieve compliance with discharge limits set by regulators and ensure tolerable burden of disease in water reuse practices, while minimizing cost and drawbacks coming from eco-toxicological impact of chemicals or energy consumption. This work addressed wastewater disinfection mostly with a process-scale perspective, conceiving predictive models to estimate process state and disturbances and how to use such models within control strategies aimed at optimal compliance with discharge limits on microbial indicators while reducing externalities. Then, part of the work is devoted to put the disinfection process in a broader perspective, considering how it impacts actual health risk in a water reuse system, conceiving and applying a

microbial risk assessment model to optimize disinfection according to a risk-based approach, which consider the process as part of an integrated system.

The PhD thesis is subdivided in four thematic chapters, each one following a scientific article-based structure. Chapters 1 and 2 are published on international scientific peer-reviewed journals, while chapters 3 to 5 are under revision or intended for publication and will be submitted soon. A schematic overview of the thesis chapters is given in Figure 1, and their content is briefly described in the following.

- *Chapter 1*: a novel approach is presented for virtual monitoring of *E. coli* concentration in WWTP secondary and tertiary effluents before disinfection, by means of a "soft-sensor" approach. "Easy-to-measure" and conventional wastewater physical and chemical parameters were explored as potential predictors of *E. coli* concentration using regression models. Data coming from three full-scale WWTP were used for model selection and testing. A neural network model turned out as the best option for soft sensor application and its potential benefits were highlighted when deployed for real time control.
- *Chapter 2*: a neural network model of full-scale UV disinfection was calibrated and tested, and used to simulate real-time control of the process to comply with agricultural water reuse limits. Results proved good accuracy of the developed model on test data and highlighted benefits coming from its deployment for real-time control of UV disinfection, which lead to an estimated energy saving of 66% with respect to business-as-usual off-line, conservative and experience-based control.
- *Chapter 3 and 4*: mechanistic model of PAA disinfection is conceived and tested (chapter 3) and deployed for Model Predictive Control (MPC) of the process (chapter 4). Lab and pilot scale studies were carried out to characterize reactor fluid dynamics *E. coli* inactivation, PAA decay and how the latter is impacted by wastewater quality. Developed model proved good performances in catching dynamic behavior of the PAA disinfection process, and its deployment in an MPC control algorithm leads to good control of effluent *E. coli* concentration.
- *Chapter 5*: QMRA was applied to an agricultural indirect water reuse system to support risk-based optimization of UV disinfection, considering accidental ingestion of soil particles by workers and crop consumption as exposure scenarios. QMRA revealed the importance of considering the whole integrated wastewater reuse system in management of disinfection

operation, which allows to set disinfectant dose considering fate and transport pathogen undergo between point of discharge and point of exposure.

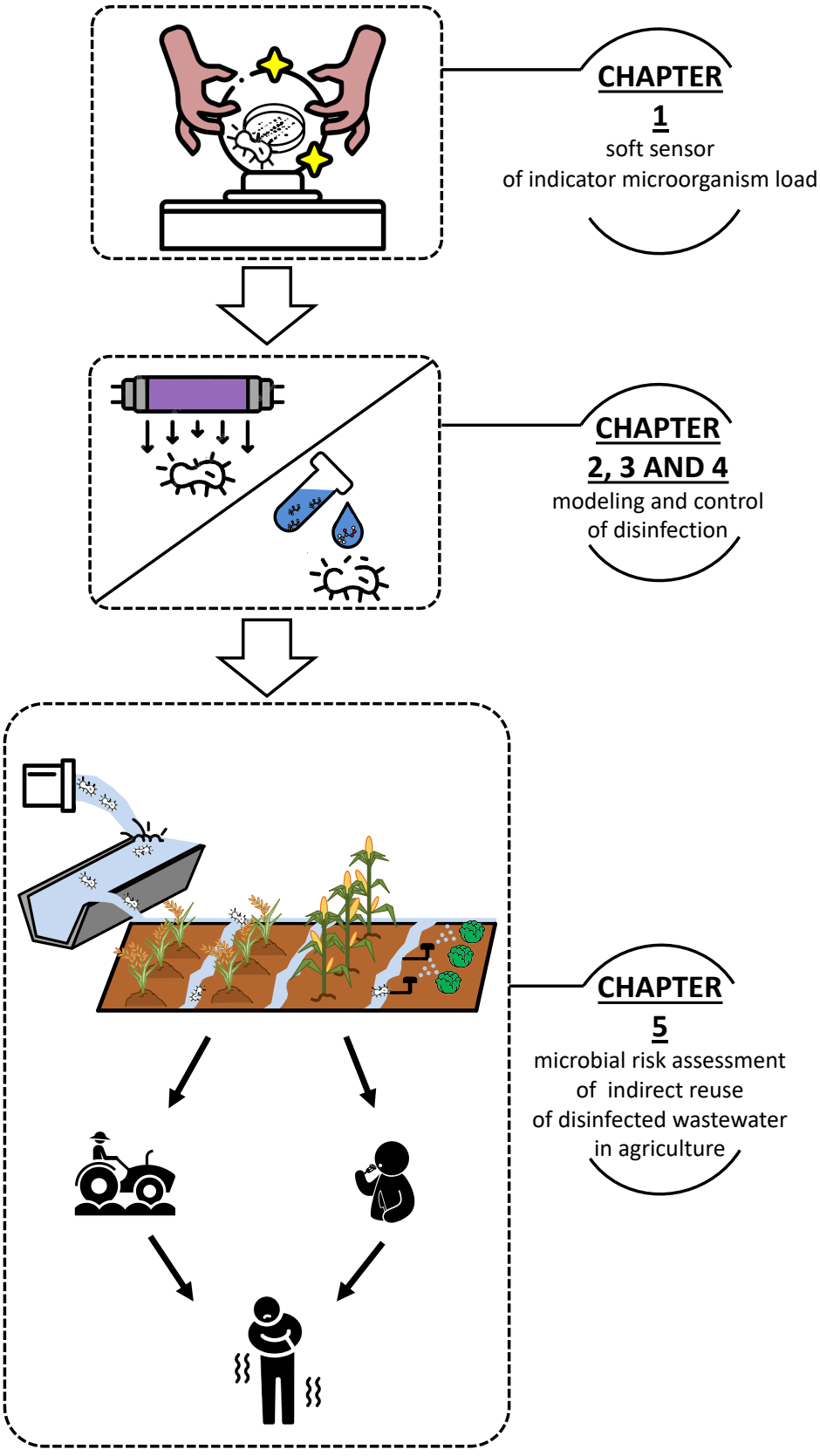

Figure 1 – Graphical concept of design of the research

Chapter 3: Soft sensor predictor of *E. coli* concentration based on conventional monitoring parameters for wastewater disinfection control

Abstract

Real-time acquisition of indicator bacteria concentration at the inlet of disinfection unit is a fundamental support to the control of chemical and ultraviolet wastewater disinfection. Culture-based enumeration methods need time-consuming laboratory analyses, which give results after several hours or days, while newest biosensors rarely provide information about specific strains and outputs are not directly comparable with regulatory limits as a consequence of measurement principles.

In this work, a novel soft sensor approach for virtual real-time monitoring of *E. coli* concentration is proposed. Conventional wastewater physical and chemical indicators (chemical oxygen demand, total nitrogen, nitrate, ammonia, total suspended solids, conductivity, pH, turbidity and absorbance at 254 nm) and flowrate were studied as potential predictors of *E. coli* concentration relying on data collected from three full-scale wastewater treatment plants. Different methods were compared: (i) linear modeling via ordinary least squares; (ii) ridge regression; (iii) principal component regression and partial least squares; (iv) non-linear modeling through artificial neural networks.

Linear soft sensors reached some degree of accuracy, but performances of the artificial neural network based models were by far superior. Sensitivity analysis allowed to prioritize the importance of each predictor and to highlight the site-specific nature of the approach, because of the site-specific nature of relationships between predictors and *E. coli* concentration. In one case study, pH and conductivity worked as good proxy variables when the occurrence of intense rain events caused sharp increases in *E. coli* concentration. Differently, in other case studies, chemical oxygen demand, total suspended solids, turbidity and absorbance at 254 nm accounted for the positive correlation between low wastewater quality and *E. coli* concentration. Moreover, sensitivity analysis of artificial neural network models highlighted the importance of interactions among predictors, contributing to 25 to 30% of the model output variance. This evidence, along with performance results, supported the idea that nonlinear families of models should be preferred in the estimation of *E. coli* concentration.

The artificial neural network based soft sensor deployment for control of peracetic acid disinfectant dosage was simulated over a realistic scenario of wastewater quality recorded by on-line sensors over 2 months. The scenario simulations highlighted the significant benefit of an *E. coli* soft sensor, which provided up to 57% of disinfectant saving.

Keywords: disinfection; soft sensor; artificial neural network; *E. coli*; peracetic acid; wastewater

This chapter has been published on "Water Research".

3.1 Introduction

Chemical and ultraviolet (UV) disinfection are two effective solutions for wastewater treatment (*inter* alia: Mezzanotte et al., 2007), while the concentration of chemical disinfectant, or UV radiation intensity, and the contact time are the two fundamental process parameters in determining pathogen inactivation. Proper control of the disinfectant dosage can compensate contact time variations and preserve disinfection efficiency, as recently demonstrated by Manoli et al. (2019). This issue is particularly important when wastewater is treated for reuse, which implies strict limits on microbial indicators in the effluent. For example, the European Parliament recently approved the new regulation for agricultural wastewater reuse (The European Parliament and the Council of the European Union, 2020), where limits on *E. coli* concentration after treatments are 10 or 100 CFU/100 mL depending on the type of crops and irrigation techniques. Nevertheless, the disinfection process is strongly affected by many sources of uncertainty, which cause high variability of the process efficiency with respect to the desired performance setpoint. Particularly, the concentration of the indicator bacteria entering disinfection is highly variable, as a result of the variability of raw wastewater characteristics, upstream treatment layout and deviation from nominal conditions of wastewater treatment plants (WWTP). Moreover, intense rain events can cause fluctuations of order of magnitudes in bacteria concentration. The first reason of inlet indicator bacteria concentration importance on disinfection is straightforward: the higher the load, the higher the required reduction, given a discharge limit to comply with. In addition, bacteria concentration can significantly affect inactivation kinetics (Haas & Kaymak, 2003; Kaymak & Haas, 2008) and, thus, has an impact on the determination of the disinfectant dosage.

Traditional disinfection modeling approaches do not address the issue of indicator bacteria load variability, and assume average, 95% percentile or maximum of available observations, as well as literature reference values, as basis for process design and operation (Ducoste et al., 2001). More recently, inlet bacteria concentration and disinfection kinetics parameters were described as multivariate random variables (Santoro et al., 2015; Ahmed et al., 2019). In these non-deterministic

frameworks, *E. coli* variability was modeled according to observed concentrations in several real WWTP effluents, accounting for stochastic dependency with inactivation kinetics parameters. These models give a realistic and reliable estimation of the indicator bacteria concentration variability upon which optimizing the disinfectant dosage. However, they are still a static description of the bacteria load, synthesizing the quantity fluctuations as a fixed probability distribution, to be estimated once for all through a site-specific data collection campaign.

Differently, providing knowledge about indicator bacteria concentration trend in time and, importantly, being able to get this information on-line, is still a challenge. Culture-based methods for estimation of indicator bacteria require time-consuming laboratory work. *E. coli*, which is one of the most widespread indicator bacteria for water and wastewater, requires at least 18 hours of incubation to be cultured (APHA, 1999). Measurement of indicator bacteria concentration is thus usually limited to mandatory analyses of the effluent at the point of discharge. The disinfection treatment is then designed and managed according to an approximate or partial knowledge about inlet bacteria concentration and wide safety factors are assumed. Sensors for real-time monitoring of bacteria concentration are under development, but mainly in the drinking water field (Skovhus & Horjis, 2018). One of the major challenges in the implementation of these sensors is that they are based on measurement principles, which are different from traditional culture-based methods. Thus, they provide outputs which are not directly comparable to regulation limits, expressed as CFU (Colony Forming Units) or MPN (Most Probable Number) values. Moreover, sensors rarely measure concentration of specific indicator bacteria strains like *E. coli* (Skovhus and Hojris, 2018).

Soft sensors for regression applications are models that exploit “easy-to-measure” variables, which can be monitored on-line at a reasonable cost, to predict target “hard-to-measure” variables (Souza et al., 2016). In recent years many studies highlighted the potential of data-driven regression models, as soft sensing tools, to predict non or hardly measurable variables in WWTPs. Particularly, many applications focus on the prediction of organic matter content and nutrients, and on treatment fault detection (Haimi et al., 2013; Corominas et al., 2018). However, wastewater disinfection applications have not been explored yet. A soft sensor could be a cost-effective solution to virtually monitor the concentration of indicator bacteria in wastewater and to optimize disinfectant dosage accordingly. Actually, the use of predictive modeling to estimate *E. coli* concentration proved promising performances in many environmental water matrices, where some chemical, biological and physical characteristics effectively worked as predictors of *E. coli* concentration (Christensen et al., 2002; Dwivedi et al., 2013, 2016; Mälzer et al., 2016; Wang et al., 2018; Rossi et al., 2020). However, a model for real-time estimation of indicator bacteria load has not been still studied for disinfection process control and optimization.

In this study, the development of a soft sensor approach to virtually monitor *E. coli* concentration at the inlet of real-scale disinfection units is presented. Conventional wastewater physical-chemical indicators, flowrate and *E. coli* were monitored in three large scale WWTPs in the Milan urban area (Italy). Both linear and nonlinear models were explored, to identify the optimal family of models and assess prediction performances. Sensitivity analysis was performed to identify negligible factors, prioritize relevant predictors and use the regression model to explore the relationship between the predictors and *E. coli* concentration. Finally, an example of the benefits of an *E. coli* soft sensor is reported, simulating its deployment on a realistic scenario for the optimization of the disinfectant dosage, focusing on the use of peracetic acid (PAA).

3.2 Materials and Methods

3.2.1 Data collection

Data collection was carried out at three large scale WWTPs in the Milan urban area: WWTP1 (1,250,000 PE), WWTP2 (600,000 PE) and WWTP3 (440,000 PE). Samples were collected at the inlet of disinfection units and analyzed for *E. coli* concentration and wastewater quality characteristics (pH, conductivity, temperature, TSS, turbidity, COD (chemical oxygen demand), UV absorbance at 254 nm (UV254), total nitrogen, ammonia, nitrate) and flowrate.

As for WWTP1, 42 samples were collected between July 2018 and March 2019, in both dry and wet weather conditions. Wastewater quality characteristics, listed in Table 1, were recorded by the disinfection unit on-line sensors (10 minutes acquisition frequency), except for UV254, which was measured in laboratory for each sample collected. As for WWTP2 and WWTP3, respectively 13 and 19 samples were collected in dry weather conditions between July 2019 and January 2020, and wastewater characteristics in Table 1 were measured in laboratory. All WWTPs were monitored at approximately weekly frequency.

Three datasets were created from the collected data, combining monitoring data from the three WWTPs, as detailed in Table 1. Dataset 1 comprises only data from WWTP1, while dataset 2 is composed grouping data from WWTP2 and WWTP3. In fact, WWTP1 is worth to be studied independently, being the only case where also rainy conditions occurred and wastewater quality was monitored with commercial on-line sensors. Finally, an additional dataset (dataset 3) combining all available data was considered, in order to explore potential general relationships between predictors and *E. coli* concentration. Differences in terms of dataset variables are due to the fact that total nitrogen and nitrate ion were monitored only in WWTP2 and WWTP3: for each dataset, only variables which were monitored in all the samples were kept in the modeling step.

Table 1 – Datasets used for soft sensor development: predictor variables, number of data available and source WWTPs.

Predictors	Dataset 1	Dataset 2	Dataset 3
pH	•	•	•
Conductivity	•	•	•
Temperature	•		
Turbidity	•	•	•
TSS	•	•	•
COD	•	•	•
UV absorbance	•	•	•
N		•	
NO_3^-		•	
NH_4^+	•	•	•
Flow rate	•		
Dataset size	42	32	74
Calibration subset size	33	26	62
Test subset size	9	6	12
Source	WWTP1	WWTP2+WWTP3	WWTP1+WWTP2+WWTP3

3.2.2 Description of WWTPs under study

The three WWTPs under study receive municipal wastewater from similar drainage basins, with a limited contribution of industrial wastewater. All the WWTPs are based on activated sludge processes for carbon and nitrogen removal. WWTP1 has a pre-denitrification stage followed by nitrification. Phosphorous is then removed by chemical precipitation with ferric chloride and removal of suspended solids is refined by sand filtration. The last stage is UV disinfection.

WWTP2 stream is firstly treated in a primary sedimentation stage. Biological treatment is made of a sequence of pre-denitrification, nitrification and post-denitrification, followed by chemical removal of phosphorous by aluminium chloride and disc filtration. Disinfection is carried out by sodium hypochlorite dosage.

WWTP3 has a primary sedimentation stage, followed by a nitrification and phosphorous removal by aluminium chloride. The stream quality is then refined by a double stage biofilter (BIOFOR®, Suez) performing nitrification, denitrification and suspended solids filtration. Disinfection is then performed by peracetic acid.

3.2.3 Sampling procedure, microbiological and chemical analysis

All samples were collected in 1-L sterile bottles, transported to the laboratory in refrigerated bags and analyzed within 12 h (conservation in refrigerated chamber at 4 °C). Putative *E. coli* were enumerated by membrane filtration method according to Standard Methods (Section 9222, APHA/AWWA/WEF, 2012), using 0.45- μm pore size cellulose nitrate membranes (Whatman) and

chromogenic agar (Microinstant® Chromogenic Coliforms Agar, Scharlau) as culture medium. Inoculated plates were incubated at 37 °C for 24 h. *E. coli* were expressed as CFU in 100 mL volume. For WWTP1, wastewater quality parameters were measured by a multiparameter monitoring station (micro::station®, S:CAN), measuring pH, conductivity, temperature (measured at sampling time, before refrigeration), COD, TSS and ammonia, and an on-line turbidimeter. For WWTP2 and WWTP3, commercial test kits were used for measurement of COD (Hach LCI500, ISO 15705), ammonia (Hach LCK303, ISO 7150-1), nitrate (Hach LCK339, EN38405 D-2) and total nitrogen (Hach LCK238, EN ISO 11905-1). TSS, turbidity and UV254 were respectively measured by 0.45- μm membrane filtration (Standard Methods, section 2540B, APHA/AWWA/WEF, 2012), portable turbidimeter (VELP Scientifica) and 1-cm optical path laboratory spectrophotometer (Hach DR6000).

3.2.4 Soft sensor development

The soft sensor for *E. coli* concentration was developed following four main steps: (i) data collection (paragraph 2.1) and pre-processing, (ii) variable selection, (iii) model selection and calibration, (iv) model testing (Fortuna et al., 2007; Souza et al., 2016). The variable selection, regression model selection and evaluation procedure were repeated for each dataset. Only 80% of data (calibration subset) was used for variable and model selection and calibration; 20% of data (test subset), randomly sampled, was used only to assess model performances in the testing step. Resulting numbers of data in each subset for the three different datasets are reported in Table 1. In detail, for dataset 1 test data corresponded to a randomly selected time window comprising 20% of the whole dataset, in order to use the test subset to both evaluating model performances and assessing the model deployment benefits over a continuous time span (10/10/2018 – 28/11/2018).

3.2.4.1 Data pre-processing

Input data were standardized to have null mean and standard deviation equal to 1, in order to ensure they have equal importance in the training process. As output data, the logarithmic transformation of *E. coli* concentration was taken, to properly manage the wide range of concentration variability, within order of magnitudes of 10^2 to 10^5 (Shu & Burn, 2004). Ammonia data were excluded from the analysis, since in most cases they resulted lower than the LOQ (Level of Quantification) and thus the parameter was not considered as a good candidate as predictor of *E. coli* concentration.

3.2.4.2 Variable selection, model selection and calibration

Four kind of linear modeling options were evaluated: (i) ordinary least squares (OLS) multiple linear regression (MLR) coupled to a best variable subset selection strategy (BVSS); (ii) ridge regression (RR) (Hoerl & Kennard, 2018); (iii) principal component regression (PCR) (Massy, 1965); (iv) partial least squares (PLS) (Wold, 1975). All the four approaches perform variable and model selection steps within the identification process itself, as following. More details about linear modeling approaches are reported in Supplementary Material (Paragraph 1).

Each one of the listed linear regression techniques requires the tuning of one hyperparameter, being the number of input variables for MLR (k), the value of the shrinking factor for RR (λ) and the number of components for PCR and PLS. Optimal hyperparameters were estimated via leave-one-out cross-validation (LOOCV), which allows estimating the expected prediction error of a specified model (Hastie et al., 2009).

As for nonlinear regression model, feed-forward backpropagation artificial neural network (ANN) was calibrated and tested on the data. ANN model was chosen as nonlinear regressor because it can reproduce any nonlinear relationship at the price of increasing its complexity (Hornik et al., 1990), i.e. increasing the number of layers and neurons per hidden layer. In this work, the "shallow" ANN model was adopted, being a network with a three-layer architecture, made by one input, one hidden and one output layer. Even in single hidden layer network, the number of weight parameters defining the model can be very high, being equal to $p(n+1)$, where n is the number of hidden neurons. In small datasets, such as in this work, the number of weights of a shallow ANN with a few hidden neurons can easily become comparable to the number of observations and predictors, and even outnumber them. The consequence is to produce a model which easily overfits calibration data and provides very poor generalization capability. For this reason, in this work the ANN model was trained using a weight decay approach, which applies the same penalty on the loss function in the minimization algorithm as the ridge regression (Equation 2.1). The integration of the penalty in the training algorithm shrinks towards zero unnecessary connections, removing the effect of irrelevant predictors and input-response relationships. Moreover, the weight decay method reduces the effect of collinearities. In this work, the ANN model had two hyperparameters to be tuned: the number of hidden neurons and the weight decay rate. In order to find the optimal hyperparameters, LOOCV was adopted. Early stopping technique was used to further avoiding overfitting and improving generalization. Validation error was monitored during each training process. Training and validation errors usually decrease together during the first training iterations. At later stages of training, the network starts to overfit the data and the validation error typically

rises. Summing weight decay and early stopping produce some redundancy in the direction of avoiding overfitting. However, the two methods tackle different issues. Weight decay shrinks irrelevant parameters, performing a sort of variable selection, even if it does not produce a sparse model. Early stopping performs regularization by looking at the validation data, which are an additional information with respect to the training subset. ANN models with up to 15 neurons were considered in the LOOCV process, representing the inner loop in the whole model selection process. An outer LOOCV loop was performed for different values of the weight decay rate, ranging between 0.0001 and 1. In each iteration of the LOOCV loop, ANN models were trained by means of back-propagation algorithm, based on gradient descent for mean squared error minimization. Details about the early stopping algorithm and other ANN training settings are reported in detail in the Supplementary Material (Paragraph 2).

Linear regression analyses were performed on software R 3.6.3, using packages `leaps 3.1` for BVSS, `glmnet 4.0` for RR and `pls 2.7-2` for PLS and PCR. Training and CV of ANN models were performed on Python 3.3.6, using PyTorch 1.3.1 and ScikitLearn 0.22.1 libraries.

3.2.4.3 Model testing

All the identified regression models were evaluated on the test subset of the corresponding dataset. Three performance metrics were used, being the Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE) and the Coefficient of Determination (R^2).

MAE and MAPE were chosen because of their easy interpretability. Differently from widely used performance metrics, like mean squared error (MSE) and root mean squared error (RMSE), MAE and MAPE are not influenced by squares and square roots, which emphasize the impact of high errors, improving model training but preventing easy and direct interpretation. For the same reason, MAE and MAPE were computed using *E. coli* concentration as CFU/100 mL. Differently, R^2 was computed over logarithmic *E. coli* concentration, since R^2 values depend on the sum of squares and can then overweight high values.

3.2.5 Sensitivity analysis

Sensitivity Analysis (SA) of identified regression models was performed to prioritize the importance of relevant predictors.

Variance-based sensitivity analysis (VSA) was chosen for predictor prioritization. VSA is an "all-at-a-time" global SA method, where output variations are obtained by varying all the inputs simultaneously, exploring their entire variability space (Saltelli et al., 2006; Pianosi et al., 2016).

The contribution of the input X_i on the output Y variability is accounted by the term S_i (first order effect), defined as:

$$S_i = \frac{VAR[E[Y|X_i]]}{VAR[Y]} \quad \text{(Eq. 2.2)}$$

The larger S_i, the larger the contribution of X_i to the output Y. When the inputs are uncorrelated, the total variance $VAR[Y]$ is the sum of the contributions given by the variation of all the single inputs X_i plus the contribution given by the interaction among inputs. Conversely, if the inputs are correlated, the sum of the variance contributions given by all the inputs might be higher than $VAR[Y]$, because the conditional variance also carries over the effects of other predictors that could be positively or negatively correlated to X_i (Saltelli et al., 2006).

A critical step in VSA is the right definition of the variability space and probability distribution for the inputs, which can significantly affect the SA results. In this work, input values for model simulations in SA were sampled from independent uniform distributions between the minimum and maximum observed values of each input, which is the most common and simple approach (Pianosi et al., 2016). The correlation among predictors was not considered in this step, since the aim of the analysis was primarily to understand how the models behave, "opening" the black-box, and then to prioritize the predictors after the model selection step. Under the assumption of uncorrelation, S_i ranges between 0 and 1. Thanks to this property, results from SA application on linear models and ANN (Marseguerra et al., 2003) can be directly compared and the effect of nonlinearity and interactions can be studied. Under the hypothesis of uncorrelation of inputs, the contributions of all the interactions among the inputs can be estimated as:

$$S_{int} = 1 - \sum_{i=1}^{p} S_i \quad \text{(Eq. 2.3)}$$

3.2.6 Simulation of soft sensor deployment

In order to highlight the benefits of a soft sensor for virtual *E. coli* monitoring and disinfection process control, *E. coli* concentration in WWTP1 undisinfected effluent was simulated by the ANN-based soft sensor model derived from regression on dataset 1, during the time span comprising test subset data (paragraph 2.1). UV254 was assumed as constant. Estimated *E. coli* concentrations were used to compute optimal PAA dosage, relying on site-specific experimental results (see Supplementary Material, Paragraph 4) about disinfectant decay and inactivation in WWTP1 wastewater entering full-scale disinfection unit. Since no immediate oxidative demand was observed in the data (Rossi et al., 2007), PAA concentration needed to comply with a selected discharge limit was calculated assuming a first order kinetic model for PAA decay. The model described by Domínguez Henao et al. (2018) was used for *E. coli* inactivation. Assuming the kinetic

parameters as constant over time, Equation 2.6 was used to compute the required exposure dose, while Equation 2.5, being the integral over time of Equation 2.4, was used to compute the PAA dosage needed to guarantee that dose. Results about the calibration of PAA kinetics are reported in detail in the Supplementary Material (Paragraph 4).

$$PAA(t) = PAA_0 e^{-kt} \quad \text{(Eq. 2.4)}$$

$$D(t) = \frac{PAA_0}{k}(1 - e^{-kt}) \quad \text{(Eq. 2.5)}$$

$$\frac{\ln(N)}{\ln(N_0)} = \frac{-k'D^n}{1-e^{h-D}} \quad \text{(Eq. 2.6)}$$

where t is time, k is the first order decay rate of PAA, PAA_0 is the PAA initial concentration, D is the dose, N is *E. coli* concentration at time t, N_0 is *E. coli* initial concentration and k', n and h are the inactivation kinetic parameters. t was assumed as equal to 30 minutes, being a typical hydraulic residence time for chemical disinfection. The ANN-based soft sensor model used in this scenario was re-calibrated (Haupt et al., 2008) on the calibration subset of dataset 1, according to two approaches. Firstly, a single ANN model was calibrated. Moreover, an ensemble of 100 ANN models (ANNe) was calibrated, according to the "bootstrap aggregation", or "bagging", approach (Breiman, 1996). Bootstrap sampling consists in randomly drawing new datasets from the original calibration data, each sample being a random sample with replacement the same size as the original dataset. In some of the re-sampled datasets some observations can appear more than once, while in other cases one or more observations can be absent. One model is fitted for each of the bootstrap samples. The probability distribution of model parameters, due to the modeling uncertainty, was approximated by the observed distribution of the parameters derived from each model training. Similarly, given a set of observations of the predictors, the predictions given by ANNe provided both a point estimate, by averaging the predictions, and a confidence interval, by taking the observed percentiles of the predictions (Efron & Hastie, 2016).

3.3 Results and Discussion

3.3.1 Wastewater characterization

Summary statistics of monitored wastewater physical and chemical characteristics and weather conditions during sample collection are reported in Table 2. Rainfall data close to WWTP1 location during the data collection campaign are reported in Supplementary Material (Figure S9). All the three monitored WWTPs have good quality effluents, both in terms of solids, organic and nitrogen (where monitored) content. One straightforward reason was the presence of tertiary treatments for quality refinement. In fact, mean effluent constituents are close to typical literature values of

tertiary effluents after biological treatment for carbon and nitrogen removal and filtration for residual solids removal (Metcalf & Eddy et al., 2013). Moreover, raw wastewater streams in the Milan area are typically significantly diluted, due to integration of some canals in the sewer network and frequent discharge of groundwater in the sewer to control water table level. *E. coli* concentration was the single parameter deviating in some cases from typical levels of a tertiary effluent, peaking 57000 CFU/100mL. Such high concentrations occurred during rain events and they could be explained by possible resuspension phenomena taking place in the sewer, resulting in higher *E. coli* concentration at the inlet of the WWTP, or to releases from sand filtration, whose efficiency could temporarily drop due to sharp increase of flow rate. Since data-driven regression approaches were used to develop the soft sensor model for each dataset, it is important to stress the fact that model validity is limited to observations ranges of variability which were recorded during the collection campaign. Limitations could be related to period of the day (e.g. night hours), the week (e.g. weekend) or the year (April, May and June, in this work) which could be not well represented by collected datasets.

Table 2 – Physical/chemical characteristics of sampled wastewater streams

Parameter	**Unit**	**WWTP1**	**WWTP2**	**WWTP3**
pH	[-]	6.84±0.13 (6.64; 7.03)	7.31±0.27 (6.94; 7.95)	7.79±0.32 (7.33; 8.77)
Conductivity	[µS cm^{-2}]	777.52±89.6 (599.4; 865)	807.23±26.38 (735; 836)	604±108 (458; 800)
Temperature	[C°]	20.2±2.9 (16.4; 24.5)	-	-
Turbidity	[NTU]	1.96±0.84 (1.21; 3.98)	3.09±1.34 (1.44; 6.35)	4.95±1.92 (2.25; 9.27)
Total suspended solids	[mg L^{-1}]	2.8±2.5 (0.1; 8.8)	6.11±4.49 (0.93; 16.4)	8.81±8.32 (2.95; 29)
Chemical oxygen demand	[mg L^{-1}]	9.96±4.56 (4.71; 20.50)	15.85±1.71 (13.5;18.9)	29.47±14.51 (9.59; 52.50)
UV absorbance at 254 nm	[-]	0.071±0.018 (0.04; 0.115)	0.121±0.022 (0.099; 0.171)	0.080±0.025 (0.051; 0.150)
Total nitrogen	[mg L^{-1}]	-	8.45±2.68 (4.19; 12.8)	10.00±3.31 (6.37;16.80)
Nitrates	[mg L^{-1}]	-	6.24±2.53 (2.829; 12.06)	23.54±11.74 (12.80; 41.50)
E. coli	[CFU/100mL]	7892±11888 (1010; 57000)	27000±12000 (7900; 44000)	177377±230441 (3000; 730000)
Flow rate	[m^3 s^{-1}]	1.5 ±0.59 (0.72; 3.93)	-	-
Rainfall	[mm]	0.5±0.8 (0.2; 9.0)	-	-

3.3.2 Model identification and performance evaluation

The model selection and calibration step aimed at identifying the best family of models and predictors. Regularization techniques used in this work aims at removing non-relevant predictors or

shrinking their coefficients, thus reducing model complexity and improving the model generalization capability. Both linear and nonlinear regression analyses were performed, in order to understand to what extent nonlinear dependencies and interactions are important in estimating input-output relationships. Since there is no model or framework in literature to estimate inlet bacteria concentration in disinfection, regression models here developed were compared to two "baseline" models (BM) which are considered as representative of the general approach in disinfection design and operation: (i) mean of the data on the calibration subset (BM1), (ii) 95th percentile of the calibration subset (BM2). The performance metrics of BM1 on the test subset was considered as the "base error rate" (Hastie et al., 2009), since observed mean can be seen as the simplest possible model, while BM2 was included in the comparison as an example of more precautionary alternative to the solution here proposed.

Test performances of all the regression options for each dataset are reported in Table 3. Details about the calibration and cross-validation results are reported in the Supplementary Material (Paragraph 3). Overall, nonlinear regression based on ANN shows the best performances and a significant improvement with respect to BM1 and linear models, suggesting that nonlinearities and interactions are fundamental to get satisfying performances from a soft sensor for *E. coli*. Regarding BM2, as shown in Table 3, the price of guaranteeing a wide safety factor is a huge average error in the test cases. Differently from linear models, the ANN model is hardly interpretable via direct inspection of parameters (weights).

Table 3 – R^2, MAE (mean absolute error) and MAPE (mean absolute percentage error) computed on test data subsets for the developed soft sensor regressors.

Model	Dataset 1			Dataset 2			Dataset 3		
	R^2	MAE	MAPE	R^2	MAE	MAPE	R^2	MAE	MAPE
BM1 - base error rate	-	7728	55%	-	193130	290%	-	29894	198%
BM2	-	22133	495%	-	291983	3300%	-	55123	1820%
MLR	0.25	7420	40%	0.39	142420	215%	0.4	22633	103%
RR	0.31	7289	42%	0.22	161827	261%	0.32	23515	177%
PCR	0.41	6980	40%	0.24	152422	206%	0.30	22765	195%
PLS	0.41	6995	41%	0.18	155602	253%	0.33	22659	185%
ANN	0.60	2337	29%	0.71	114515	130%	0.79	11686	47%

*f*table 4

3.3.3 Sensitivity analysis: discussion on physical and chemical parameters relevance

Sensitivity analysis (SA) is of fundamental importance for model developers and users to understand key features of the model and exploit the model at its highest potential. This kind of analysis is particularly important in case of high mathematical complexity models, like ANNs, whose parameters lack of direct interpretability and physical significance (Olden & Jackson, 2002).

Actually, data-driven regression models capture input-output relationships as they perceive them in the available training data. After model identification, it is important in engineering applications to interpret the predictive relationships derived from the data and to compare the regression analysis results with *a priori* knowledge about phenomena. The use of VSA presents some advantages in inspecting the model, as briefly detailed in the Supplementary Material (Paragraph 7).

Values of S_i from the VSA are reported in Figure 1, computed both for the RR model, as an example of a linear model, and the ANN model, for the three studied datasets. In many cases the index is almost null, as a result of the shrinking effect on the model parameters produced by the penalty on the loss function imposed both by RR and weight decay in ANN training. Results about dataset 1 show that pH is the most relevant factor in determining the variance of the output, both for the linear and the nonlinear model cases (30-40%). In the nonlinear case, conductivity resulted also as an important factor, with about 10% contribution to the total variance. The relationship with *E. coli* concentration is negative both for pH and conductivity. A possible interpretation is that the most significant increase in *E. coli* concentration occurs during intense rain events, when the contribution of stormwater lowers pH and conductivity. Stormwater can actually determine sharp increase in *E. coli* concentration, due to urban runoff and sediment re-suspension in the sewer (McCarthy, 2009; Hathaway & Hunt, 2011; McKergow et al., 2010). The use of pH and conductivity as proxies of the incoming of stormwater, instead of directly using the rainfall data as a predictor, avoids the critical issues of modeling the delay between the occurrence of rain events and the incoming of stormwater in the WWTP, which implies the description of the corrivation phenomena and the choice of optimal rainfall monitoring sites. Moreover, acquisition of pH and conductivity data from on-line sensors is a much more practical solution than implementing the real-time transmission of rainfall data from monitoring sites. The WWTP1 case study is then a clear example of how a soft sensor can exploit available monitoring data in order to catch proxy relationships which work as surrogates of the real physical cause-effect laws of the target phenomena: changes in conductivity or pH are not the physical cause of *E. coli* increase, but they are indirectly correlated to that. The importance of pH and conductivity stands only for this dataset. In the other two cases, their contribution to the output variance is almost negligible. Since no rain events occurred during data collection, pH and conductivity lose their role of proxies of wet weather conditions. Unexpectedly, temperature did not result as an important predictor for the case of dataset 1: probably, temperature in the effluent varied in a too narrow range to determine significant changes in *E. coli* concentration. However, since a relationship between temperature and bacteria vitality exists, the role of temperature should be further studied.

In models trained over dataset 2, most of the output variance (about 50%) is caused by turbidity. Significant correlations between turbidity and *E. coli* concentration were found many times in surface waters (Christensen et al., 2002; Jamieson et al., 2005; Money et al., 2009; Huey & Meyer, 2010; Wu et al., 2011), which can be explained by the tendency of *E. coli* to attach to particles (Hipsey et al., 2006; Dickenson & Sansalone, 2012; Liu et al., 2013; Jiang et al., 2015). Association of coliforms on suspended solids in wastewater is still poorly explored, but it has been highlighted in some studies (Loge et al., 2002; Falsanisi et al., 2008; McKergow & Davies-Colley, 2010; Chahal et al., 2016; Carré et al., 2018). It is then highly probable that increases in *E. coli* concentration in WWTP2 and WWTP3 were determined by a higher contribution of the particle associated fraction in the influent, due to temporary decrease of upstream treatments efficiencies or higher suspended solids loads to the plant.

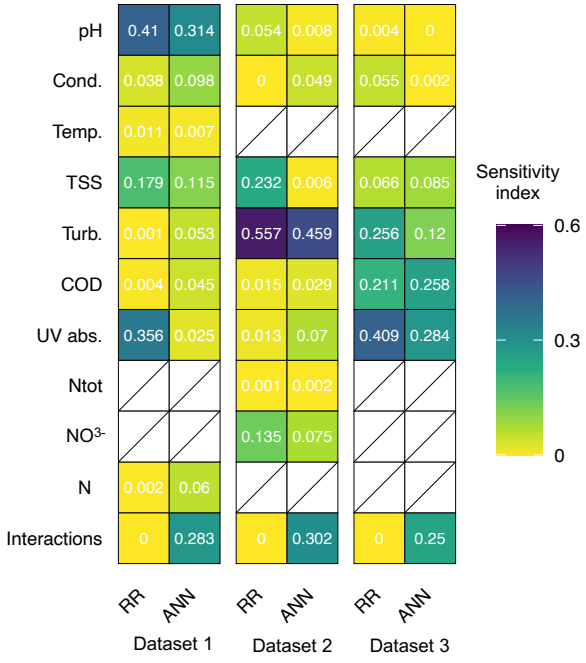

Figure 1 – Results of sensitivity analysis for RR and ANN models.

In the case of dataset 3, models were trained and validated on data coming from all the three WWTPs considered in this study. Thus, corresponding sensitivity analysis results are better candidates than the dataset 1 and 2 case to describe general relationships between physical and chemical characteristics of wastewater and *E. coli* concentration. For dataset 3, UV254, COD, turbidity and TSS resulted as most important predictors. For all the four predictors, the correlation with *E. coli* is positive, thus in this general case, the soft sensor captures the increase in *E. coli* concentration that is correlated with a global worsening in wastewater quality in terms of organic content and suspended solids. Dependence of coliforms concentration on treatment efficiency of organic matter and solids was already highlighted by George et al., 2002 and Koivunen et al., 2003.

The role of TSS and turbidity can be explained as for dataset 2, but in these cases both the two aggregated indicators of suspended solids turned out to be important, which can be explained by the difference in their measuring principles. Turbidity, which is an optical method, can be affected by particle size (Gregory, 2009; He & Nan, 2012; Yao et al., 2014), with larger particles usually contributing more to the overall nephelometric turbidity units (NTUs). Since particle size was pointed as a factor in determining particle association of coliform, with bacteria typically attaching on larger particles (Falsanisi et al., 2008; Chahal et al., 2016), both turbidity and TSS were probably kept by the model as predictors, in order to account both for total suspended solids mass and particle size, as well as their interaction. The importance of COD was likely due in part to the fact that a fraction of the residual organic matter in the effluent was in the form of suspended solids. However, the almost equal relevance of ABS254 indicate that soluble organic matter is correlated to *E. coli* concentration as well. A candidate reason behind this evidence is that lower efficiencies of the activated sludge treatments in removing soluble organic carbon can be due to lower biological activities and thus weaker "ecological barriers" for *E. coli*. It was in fact highlighted in literature by long time how predation, especially by protozoa, is the main mechanism determining abatement of fecal bacteria in biological treatments (van Der Drift et al., 1977; Loge et al., 2002). Non-optimal conditions of activated sludge may also cause the formation of smaller and less settling flocs, thus more difficultly removable by secondary settlers and filters. This could be another reason why increase in SST and turbidity work as predictors of increases in *E. coli* concentration.

Overall, there are some big discrepancies among the results from the three different datasets that worth to be discussed. First of all, as already mentioned, dataset 1 leads to models which are completely different from the other two cases, because in WWTP1 occurrence of intense rain events resulted as the only cause of sharp increases in *E. coli* concentration, determining variations of one order of magnitude. Results coming from dataset 1 are then really site-specific and no general interpretation of the phenomena can be inferred. The different ranges of variability of *E. coli* concentration in the different datasets, highlighted in Figure 2 and 3, have to be considered as well. Data from dataset 1 are mainly comprised between 1,000 and 10,000 CFU/100 mL, with a minor portion of data regarding wet weather conditions, which reach about 60,000 CFU/100 mL. Differently, data from dataset 2 range between 10,000 to 500,000 CFU/100 mL. Similarly, COD, TSS, UV254 and turbidity data vary in different ranges in the three datasets. The three datasets complement each other and thus their combination allows to better explore their space of variability and consequently reveal more relationships with *E. coli* concentration. This is one of the reasons why predictors ranking changed completely from dataset 1 to dataset 2: in the latter case the wider distribution of data in their variability space allows to the regression analysis to catch the role of

turbidity in predicting *E. coli* concentration. Since dataset 3 is the sum of all the collected data from all the three studied WWTPs, it is the one having the most complete picture on *E. coli* concentration and wastewater quality variability and thus, in addition to turbidity, it can capture the role of additional predictors, being UV254, COD and TSS.

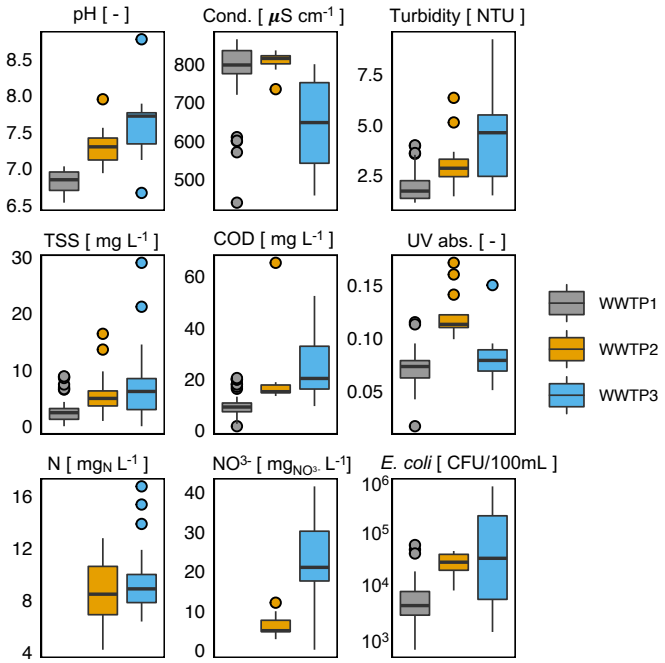

Figure 2 – Boxplots of data collected about wastewater characteristics in the three WWTP (wastewater treatment plant) under

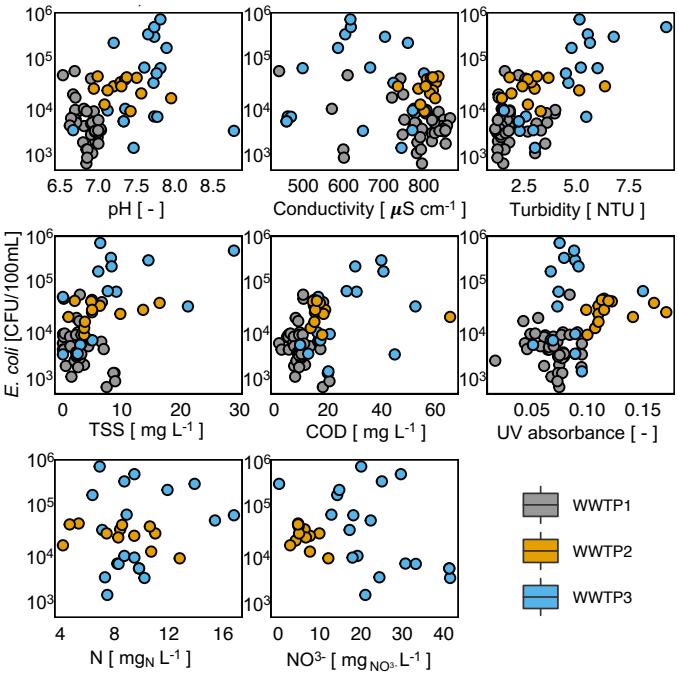

Figure 3 – Observed trend of *E. coli* concentration vs. the monitored wastewater characteristics.

SA on ANN-based soft sensors revealed that interactions among predictors are responsible of an important portion of the total variance (25-30%). This result, together with the significant increase

in performances with respect to the linear models, highlights that the relationship between *E. coli* concentration and wastewater quality characteristics is mainly nonlinear and thus a family of nonlinear models, like ANN, is the best candidate for soft sensor.

In conclusion, the coupling of model selection techniques and SA, revealed how wastewater characteristics can be screened and prioritized in order to understand which are essential to build an effective predictive model as an *E. coli* soft sensor.

3.3.4 Example of model deployment over a scenario

Figure 4 reports predictions during the time window of the ANN-based soft sensor calibrated on dataset 1. The single ANN model predictions approximate well almost all the test measurements. Importantly, the soft sensor catches the increase in *E. coli* concentration occurring around 01/11/2018, which coincided with intense rain events. Even if the model cannot explicitly discriminate between dry and wet weather conditions, it captures the sharp increase in *E. coli* concentration thanks to the wastewater quality variables which are real-time monitored and can work as proxies of a higher contribution of stormwater to the overall flow rate. Thus, the model testing support the interpretation of SA results previously reported.

Predictions from bootstrap averaging of the ANNe are also reported in Figure 4. Average predictions are similar to the single ANN model, except for the wet weather condition cases. In this period, average predictions of the ANNe are significantly lower than the single ANN, leading to a dramatical underestimation of *E. coli* in a critical event. However, the bagging approach has the advantage of producing an estimate of the confidence interval of the prediction: the 95% confidence interval always contains all the observed concentrations of the test subset. The upper bound of the ANNe is then a valuable reference to describe *E. coli* concentration when the goal is the optimization of disinfectant dosage, since it both provides an adequate safety factor and catches the main fluctuation in *E. coli* load occurring in the considered time window.

Table 4 reports PAA consumption that would be required during the whole time window considering the different approaches for *E. coli* concentration estimation. Consumptions data were computed for discharge limits equal to 5,000 and 10 CFU/100 mL, which are respectively the recommended limit for surface water discharge and the mandatory limit for agricultural reuse in Italy. The BM1 approach was totally inadequate for the discharge limit into surface water, since the average *E. coli* concentration observed in the calibration subset was lower than the discharge limit itself. Considering *E. coli* as a constant equal to the observed average leads to a critical underestimation which even leads to conclude that no PAA dosage would be needed. Considering all the other approaches, as expected the BM2 brings to the highest PAA consumption,

guaranteeing a good safety factor, since only the most critical *E. coli* concentration is underestimated. When the single ANN model is used, a PAA mass saving is obtained, being, respectively, 57% for the discharge limit in surface water and 55% for the reuse limit. Moreover, even if *E. coli* concentration is often underestimated, the error is quite small (Table 3). Lastly, when the ANNe is used, considering the upper bound of the 95% confidence interval, a reduction in PAA consumption of 12% and 25% and a wide enough safety factor, in order to never underestimate *E. coli* concentration, were obtained.

Table 4 – Mass of consumed PAA (peracetic acid) and number of under-estimation cases for every modeling approach estimating *E. coli* concentration entering the disinfection unit (N_0).

	PAA (kg)		$N_0 > \hat{N}_0$ cases
Discharge limit (*E. coli*, CFU/100 mL)	5,000	10	
BM1	0	1,084	6/9
BM2	798	4,589	1/9
ANN	140	2,050	7/9
ANNe (97.5%)	703	3,433	0/9

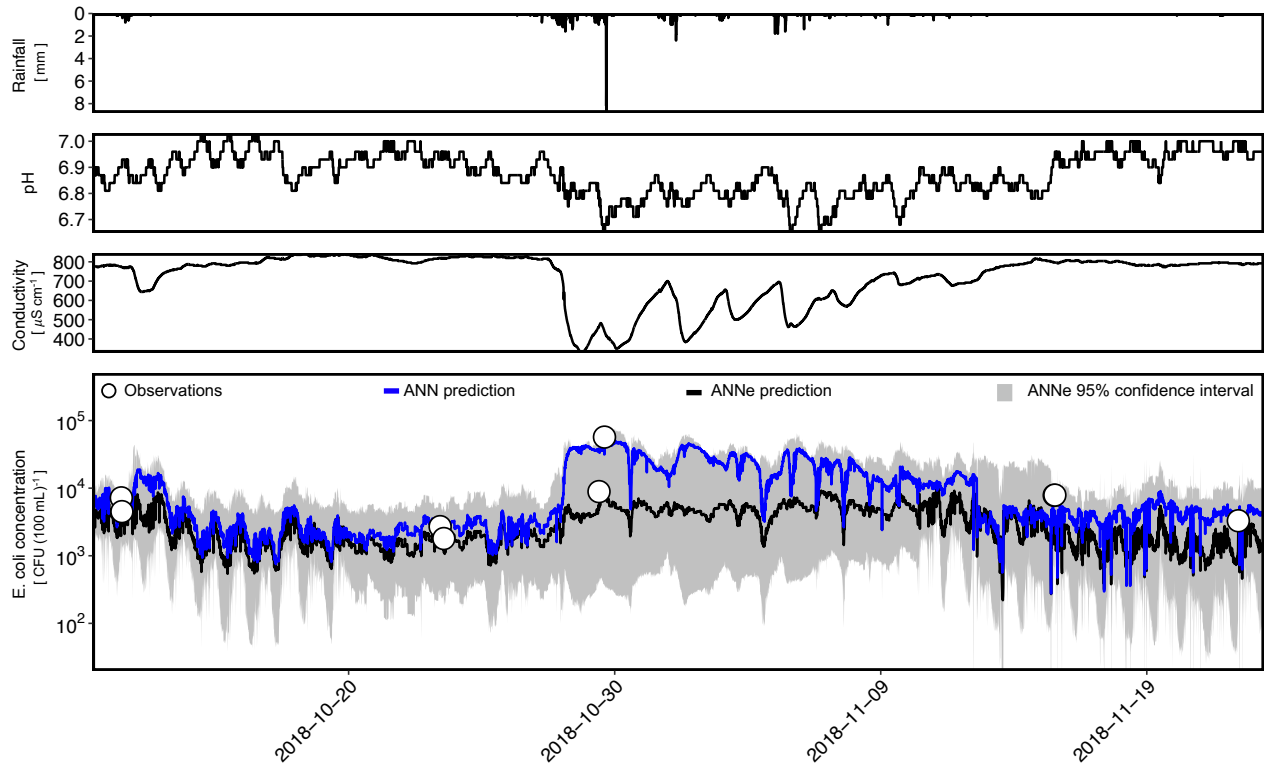

Figure 4 – Simulation of *E. coli* concentration entering the disinfection unit. Trends in rainfall, pH and conductivity during the scenario are reported. *E. coli* test observations are reported as white dots. Single ANN (artificial neural network) model prediction (blue line) and ANNe (artificial neural network ensemble) predictions (black line and grey shaded area) are shown.

3.4 Conclusions

In this study an innovative soft sensor based on "black-box" regression models is proposed for real-time virtual monitoring of *E. coli* concentration at the inlet of full-scale disinfection units. Relying on data from three WWTPs, soft sensor regressors using different physical and chemical characteristics as predictors were calibrated and gave satisfying performances. Performances of the ANN-based soft sensors were significantly higher than any linear model tested, highlighting the importance of nonlinearities and interactions in the relationships between wastewater physical and chemical characteristics and *E. coli* concentration.

Based on the results of the sensitivity analysis, in one case study, the soft sensor was able to capture *E. coli* concentration variations due to intense rain events by exploiting pH and conductivity as proxy variables of the presence of stormwater in the flow rate. Differently, when data coming from all the case studies are considered, variables describing a general worsening of the upstream treatment efficiency, being COD, UV absorbance, TSS and turbidity, resulted as the driving predictors. Discrepancies of the results among the datasets underlined how this kind of black-box regressor can have strongly site-specific validity when data from just one case study are considered. The combinations of data from three WWTPs effluents allows to develop a model which was likely more generalizable. Then, this work could be additionally developed by collecting a bigger dataset with data coming from a higher number of WWTPs, which could more robustly substantiate general conclusions on the physical cause-effect relationships describing the phenomena.

The ANN-based soft sensor was tested over a scenario of wastewater quality and the model predictions of *E. coli* concentration were used to simulate the control of the dosage of a PAA disinfection treatment. The simulation highlighted the significant benefits of the deployment of an *E. coli* soft sensor both in terms of PAA saving (up to 57%) and process reliability.

This study highlighted the potential of the soft sensor solution to unlock a new dynamic approach in optimization and control of disinfection which accounts for the indicator bacteria load dynamics, by-passing the drawbacks of laboratory culture methods and new on-line bacteria sensors. Different regularization techniques were compared, in order to deal with a limited dataset. Establishing a monitoring routine of indicator bacteria at the inlet of WWTPs disinfection by plant managers could help in building big enough datasets to improve the soft sensor robustness before deployment. Particularly, other important parameters could be monitored and studied, such as dissolved oxygen concentration, which could affect microbial vitality. After an initial intensive data collection, the monitoring protocol should include a periodical data collection in order to perform

proper model maintenance, since black-box regression models could be biased by unexpected modifications in the underlying physical system.

References

Ahmed, Y. M., Ortiz, A. P., & Blatchley, E. R. (2019). Stochastic Evaluation of Disinfection Performance in Large-Scale Open-Channel UV Photoreactors. Journal of Environmental Engineering (United States), 145(10). https://doi.org/10.1061/(ASCE)EE.1943-7870.0001562

APHA. (1999). Standard Methods for the Examination of Water and Wastewater PREFACE TO THE TWENTIETH EDITION Standard Methods for the Examination of Water and Wastewater. 2671.

Breiman, L. (1996). Bagging Predictors, URL: https://link.springer.com/article/10.1007%2FBF00058655. Machine Learning, 24(421), 123–140. https://doi.org/10.1007/BF00058655

Carré, E., Pérot, J., Jauzein, V., & Lopez-Ferber, M. (2018). Impact of suspended particles on UV disinfection of activated-sludge effluent with the aim of reclamation. Journal of Water Process Engineering, 22(January), 87–93. https://doi.org/10.1016/j.jwpe.2018.01.016

Chahal, C., van den Akker, B., Young, F., Franco, C., Blackbeard, J., & Monis, P. (2016). Pathogen and Particle Associations in Wastewater: Significance and Implications for Treatment and Disinfection Processes. Advances in Applied Microbiology, 97. https://doi.org/http://dx.doi.org/10.1016/bs.aambs.2016.08.001

Christensen, V. G., Rasmussen, P. P., & Ziegler, A. C. (2002). Real-time water quality monitoring and regression analysis to estimate nutrient and bacteria concentrations in Kansas streams. Water Science and Technology, 45(9), 205–211. https://doi.org/10.2166/wst.2002.0240

Corominas, L., Garrido-Baserba, M., Villez, K., Olsson, G., Cortés, U., & Poch, M. (2018). Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques. Environmental Modelling and Software, 106, 89–103. https://doi.org/10.1016/j.envsoft.2017.11.023

Dickenson, J. A., & Sansalone, J. J. (2012). Distribution and disinfection of bacterial loadings associated with particulate matter fractions transported in urban wet weather flows. Water Research, 46(20), 6704–6714. https://doi.org/10.1016/j.watres.2011.12.039

Domínguez Henao, L., Cascio, M., Turolla, A., & Antonelli, M. (2018). Effect of suspended solids on peracetic acid decay and bacterial inactivation kinetics: Experimental assessment and definition of predictive models. Science of the Total Environment, 643, 936–945. https://doi.org/10.1016/j.scitotenv.2018.06.219

Ducoste, J., Carlson, K., & Bellamy, W. (2001). The integrated disinfection design framework approach to reactor hydraulics characterization. Journal of Water Supply: Research and Technology - AQUA, 50(4), 245–261. https://doi.org/10.2166/aqua.2001.0021

Dwivedi, D., Mohanty, B. P., & Lesikar, B. J. (2013). Estimating Escherichia coli loads in streams based on various physical, chemical, and biological factors. Water Resources Research, 49(5), 2896–2906. https://doi.org/10.1002/wrcr.20265

Dwivedi, D., Mohanty, B. P., & Lesikar, B. J. (2016). Impact of the Linked Surface Water-Soil Water-Groundwater System on Transport of E. coli in the Subsurface. Water, Air, and Soil Pollution, 227(9). https://doi.org/10.1007/s11270-016-3053-2

Efron, B., & Hastie, T. (2016). Computer age statistical inference: Algorithms, evidence, and data science. In Computer Age Statistical Inference: Algorithms, Evidence, and Data Science. https://doi.org/10.1017/CBO9781316576533

Falsanisi, D., Gehr, R., Liberti, L., & Notarnicola, M. (2008). Effect of suspended particles on disinfection of a physicochemical municipal wastewater with peracetic acid. Water Quality Research Journal of Canada, 43(1), 47–54. https://doi.org/10.2166/wqrj.2008.006

Fortuna, L., Graziani, S., Rizzo, A., & Xibilia, M. G. (2007). Soft Sensors for Monitoring and Control of Industrial Processes. Springer.

George, I., Crop, P., & Servais, P. (2002). Fecal coliform removal in wastewater treatment plants studied by plate counts and enzymatic methods. 36, 2607–2617.

Gregory, J. (2009). Monitoring particle aggregation processes. Advances in Colloid and Interface Science, 147–148(C), 109–123. https://doi.org/10.1016/j.cis.2008.09.003

Haas, C. N., & Kaymak, B. (2003). Effect of initial microbial density on inactivation of Giardia muris by ozone. Water Research, 37(12), 2980–2988. https://doi.org/10.1016/S0043-1354(03)00112-X

Haimi, H., Mulas, M., Corona, F., & Vahala, R. (2013). Environmental Modelling & Software Data-derived soft-sensors for biological wastewater treatment plants : An overview. Environmental Modelling and Software, 47, 88–107. https://doi.org/10.1016/j.envsoft.2013.05.009

Hathaway, J. M., & Hunt, W. F. (2011). Evaluation of first flush for indicator bacteria and total suspended solids in urban stormwater runoff. Water, Air, and Soil Pollution, 217(1–4), 135–147. https://doi.org/10.1007/s11270-010-0574-y

Haupt, S. H., Pasini, A., & Marzban, C. (2008). Artificial Intelligence Methods in the Environmental Sciences. Springer.

He, W., & Nan, J. (2012). Study on the impact of particle size distribution on turbidity in water. Desalination and Water Treatment, 41(1–3), 26–34. https://doi.org/10.1080/19443994.2012.664675

Hipsey, M. R., Brookes, J. D., Regel, R. H., Antenucci, J. P., & Burch, M. D. (2006). In situ evidence for the association of total coliforms and Escherichia coli with suspended inorganic particles in an Australian reservoir. Water, Air, and Soil Pollution, 170(1–4), 191–209. https://doi.org/10.1007/s11270-006-3010-6

Hoerl, A. E., & Kennard, R. W. (2018). American Society for Quality Ridge Regression : Biased Estimation for Nonorthogonal Problems American Society for Quality Stable URL : http://www.jstor.org/stable/1267351 Linked references are available on JSTOR for this article : Ridge Regression : Biase. 12(1), 55–67.

Hornik, K., Stinchcombe, M., & White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Networks, 3(5), 551–560. https://doi.org/10.1016/0893-6080(90)90005-6

Metcalf & Eddy, Tchobanoglous, G., Stensel, D., Tsuchihashi, R., & Burton, F. (2013). Wastewater Engineering: Treatment and Resource Recovery (5th editio). McGraw-Hill.

Jamieson, R. C., Joy, D. M., Lee, H., Kostaschuk, R., & Gordon, R. J. (2005). Resuspension of Sediment-Associated Escherichia coli in a Natural Stream . Journal of Environmental Quality, 34(2), 581–589. https://doi.org/10.2134/jeq2005.0581

Jiang, L., Chen, Y. C., Zhu, D. J., & Liu, Z. W. (2015). Faecal coliform attachment to settleable suspended sediments in fresh surface waters: Linear partition model validation and sediment concentration effects. Water Science and Technology: Water Supply, 15(4), 864–870. https://doi.org/10.2166/ws.2015.042

Kaymak, B., & Haas, C. N. (2008). Effect of initial microbial density on inactivation of Escherichia coli by monochloramine. Journal of Environmental Engineering and Science, 7(3), 237–245. https://doi.org/10.1139/S07-054

Koivunen, J., Siitonen, A., & Heinonen-tanski, H. (2003). Elimination of enteric bacteria in biological–chemical wastewater treatment and tertiary filtration units. Water Research, 37, 690–698.

Liu, G., Ling, F. Q., Magic-Knezev, A., Liu, W. T., Verberk, J. Q. J. C., & Van Dijk, J. C. (2013). Quantification and identification of particle-associated bacteria in unchlorinated drinking water from three treatment plants by cultivation-independent methods. Water Research, 47(10), 3523–3533. https://doi.org/10.1016/j.watres.2013.03.058

Loge, F. J., Emerick, R. W., Ginn, T. R., & Darby, J. L. (2002). Association of coliform bacteria with wastewater particles: Impact of operational parameters of the activated sludge process. Water Research, 36(1), 41–48. https://doi.org/10.1016/S0043-1354(01)00204-4

Mälzer, H. J., aus der Beek, T., Müller, S., & Gebhardt, J. (2016). Comparison of different model approaches for a hygiene early warning system at the lower Ruhr River, Germany. International Journal of Hygiene and Environmental Health, 219(7), 671–680. https://doi.org/10.1016/j.ijheh.2015.06.005

Manoli, K., Sarathy, S., Maffettone, R., & Santoro, D. (2019). Advanced control strategies for chemical-based disinfection processes in municipal wastewater : modelling and validation studies.

Marseguerra, M., Masini, R., Zio, E., & Cojazzi, G. (2003). Variance decomposition-based sensitivity analysis via neural networks. Reliability Engineering and System Safety, 79(2), 229–238. https://doi.org/10.1016/S0951-8320(02)00234-X

McCarthy, D. T. (2009). A traditional first flush assessment of E. coli in urban stormwater runoff. Water Science and Technology, 60(11), 2749–2757. https://doi.org/10.2166/wst.2009.374

McKergow, L. A., & Davies-Colley, R. J. (2010). Stormflow dynamics and loads of Escherichia coli in a large mixed land use catchment. Hydrological Processes. https://doi.org/10.1002/hyp

Mezzanotte, V., Antonelli, M., Citterio, S., & Nurizzo, C. (2007). Wastewater Disinfection Alternatives: Chlorine, Ozone, Peracetic Acid, and UV Light. Water Environment Research, 79(12), 2373–2379. https://doi.org/10.2175/106143007x183763

Money, E. S., Carter, G. P., & Serre, M. L. (2009). Modern space/time geostatistics using river distances: Data integration of turbidity and E. coli measurements to assess fecal contamination along the Raritan River in New Jersey. Environmental Science and Technology, 43(10), 3736–3742. https://doi.org/10.1021/es803236j

Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": A randomization approach for understanding variable contributions in artificial neural networks. Ecological Modelling, 154(1–2), 135–150. https://doi.org/10.1016/S0304-3800(02)00064-9

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. Environmental Modelling and Software, 79, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Rossi, A., Wolde, B. T., Lee, L. H., & Wu, M. (2020). Prediction of recreational water safety using Escherichia coli as an indicator: case study of the Passaic and Pompton rivers, New Jersey. Science of the Total Environment, 714, 136814. https://doi.org/10.1016/j.scitotenv.2020.136814

Rossi, S., Antonelli, M., Mezzanotte, V., & Nurizzo, C. (2007). Peracetic Acid Disinfection: A Feasible Alternative to Wastewater Chlorination. Water Environment Research, 79(4), 341–350. https://doi.org/10.2175/106143006x101953

Saltelli, A., Ratto, M., Tarantola, S., & Campolongo, F. (2006). Sensitivity analysis practices: Strategies for model-based inference. Reliability Engineering and System Safety, 91(10–11), 1109–1125. https://doi.org/10.1016/j.ress.2005.11.014

Santoro, D., Crapulli, F., Raisee, M., Raspa, G., & Haas, C. N. (2015). Nondeterministic Computational Fluid Dynamics Modeling of Escherichia coli Inactivation by Peracetic Acid in Municipal Wastewater Contact Tanks. Environmental Science and Technology, 49(12), 7265–7275. https://doi.org/10.1021/es5059742

Shu, C., & Burn, D. H. (2004). Artificial neural network ensembles and their application in pooled flood frequency analysis. Water Resources Research, 40(9), 1–10. https://doi.org/10.1029/2003WR002816

Skovhus, T. L., & Horjis, B. (2018). Microbiological Sensors for the Drinking Water Industry. IWA Publishing. https://doi.org/10.2166/9781780408699

Souza, F. A. A., Araújo, R., & Mendes, J. (2016). Review of soft sensor methods for regression applications. Chemometrics and Intelligent Laboratory Systems, 152, 69–79. https://doi.org/10.1016/j.chemolab.2015.12.011

The European Parliament and the Council of the European Union. (2020). Regulation (EU) 2020/741 of 25 May 2020 on minimum requirements for water reuse. Official Journal of the European Union, 2019(April), L 177/32-L 177/55.

van Der Drift, C., van Seggelen, E., Stumm, C., Hol, W., & Tuinte, J. (1977). Removal of Escherichia coli in wastewater by activated sludge. Applied and Environmental Microbiology, 34(3), 315–319. https://doi.org/10.1128/aem.34.3.315-319.1977

Wang, C., Schneider, R. L., Parlange, J. Y., Dahlke, H. E., & Walter, M. T. (2018). Explaining and modeling the concentration and loading of Escherichia coli in a stream—A case study. Science of the Total Environment, 635, 1426–1435. https://doi.org/10.1016/j.scitotenv.2018.04.036

Wu, J., Rees, P., & Dorner, S. (2011). Variability of E. coli density and sources in an urban watershed. Journal of Water and Health, 9(1), 94–106. https://doi.org/10.2166/wh.2010.063

Yao, M., Nan, J., & Chen, T. (2014). Effect of particle size distribution on turbidity under various water quality levels during flocculation processes. Desalination, 354, 116–124. https://doi.org/10.1016/j.desal.2014.09.029

3.5 Supplementary Material

3.5.1 Linear model identification

3.5.1.1 Multiple Linear Regression via Ordinary Least Squares

Multiple Linear Regression (MLR) via Ordinary Least Squares (OLS) is still the most widespread linear modeling technique. Given p as the number of possible predictors, BVSS selects the subset of predictors of each size k, ranging between 0 and p, which minimizes the residual sum of squares of a MLR model using that subset as input. The optimal subset size k is the one which minimizes the prediction error (Hastie, 2009).

3.5.1.2 Ridge Regression

Ridge Regression (RR) identifies a linear model and shrinks the regression coefficients by imposing a penalty on their size. The penalized residual sum of squares minimized by RR is defined as:

$$SSE^{RR} = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \qquad \text{(Eq. 2.1)}$$

where y_i are the observations, $\hat{y}_i$ are the predicted values, N is the size of the sample and β_j are model parameters. RR shrinks continuously towards zero coefficients of irrelevant predictors, minimizing the impact of correlated predictors and instead promoting a "grouping effect" (Zou & Hastie, 2005; Hastie et al., 2009; Zhou, 2013).

3.5.1.3 Principal Component Regression and Partial Least Squares

Principal Component Regression (PCR) and Partial Least Squares (PLS) are two dimensionality reduction methods which replace the original predictors with a reduced number of linear combinations of them. In the case of PCR, these linear combinations are the principal components (PCs) of the input variables. PCs represent a new set of input variables, which are orthogonal and tend to capture most of the input data variance in a number of components lower than the input number, due to the replacement of the original correlated input variables.

Similarly, the PLS model has a reduced number of components, but the response variable is involved in their identification. PLS uncorrelated components are both a good representation of input data variability, and the most correlated to the response (Hastie et al., 2009).

3.5.1.4 Additional comments on linear modeling methods and results

RR, PCR and PLS have the advantage of reducing multicollinearity problems, i.e. the impact of correlation among predictors, which negatively affect coefficient estimates and are typical in case of wastewater physical-chemical properties.

For each dataset, linear regressors are associated to lower MAE and MAPE with respect to the BM1 case. There is no dramatical difference among linear model performances. However, for each dataset, a different linear modeling technique performs the best. Based on MAPE, best linear models for dataset 1, 2 and 3 are, respectively, RR, PCR and BVSS-MLR. This result underlines the importance of testing different modeling approaches for model selection and calibration, since none of them can be *a priori* claimed to be the best one (Frank & Friedman, 1993).

Parameters of all the identified linear models are reported in Table 3. Thanks to the different regularization techniques, some predictor coefficients are exactly or close to 0, corresponding to the variables which are less effective as predictors of *E. coli* concentration.

Table S1 – Coefficients of linear regression models, being OLS (ordinary least squares), RR (ridge regression), PCR (principal components regression) and PLS (partial least squares). Results of hyperparameters tuning are reported in parenthesis, where "N" stands for the number of predictors or principal components.

	Dataset1				Dataset 2				Dataset 3			
Predictors	OLS** (N= 3)	RR (λ=0.076)	PCR (N=5)	PLS (N=3)	OLS** (N=2)	RR (λ=0.095)	PCR (N=5)	PLS (N=4)	OLS** (N=2)	RR (λ=0.139)	PCR (N=5)	PLS (N=4)
pH	0.22*	0.160	-0.109	-0.168	0	0.043	0.053	0.080	0	0.048	0.080	0.045
Conductivity	0	-0.038	-0.043	-0.037	0	-0.002	0.020	-0.017	0	-0.017	0.010	-0.014
Temperature	0	0.031	0.064	0.047	-	-	-	-	-	-	-	-
Turbidity	0	-0.008	-0.042	-0.019	0.459*	0.350	0.493	0.457	0.400*	0.283	0.366	0.386
TSS	-0.14*	-0.111	-0.085	-0.128	0	0.04	-0.038	-0.013	0	0.047	0.058	0.017
COD	0	-0.015	-0.052	-0.048	0	0.017	0.014	-0.003	0	0.020	-0.054	-0.010
UV abs.	0.13*	0.103	0.178	0.175	0	-0.005	-0.023	-0.038	0.163*	0.131	0.150	0.160
N	-	-	-	-	0	0.024	-0.009	0.007	-	-	-	-
NO_3^-	-	-	-	-	-0.291*	-0.261	-0.312	-0.370	-	-	-	-
Flow rate	0	0.011	-0.063	0.008	-	-	-	-	-	-	-	-
Intercept	3.596	3.602	3.604	3.604	4.444	4.437	4.396	4.396	3.952	3.951	3.946	3.946

* p-value related to t-tests on model coefficient statistical significance was <0.05. In case of coefficient equal to 0, the predictor was excluded by the Best Variable Subset Selection algorithm and thus from the Ordinary Least Squares calculation.
**p-value related to F-test on model statistical significance was <0.05

3.5.2 ANN training, cross-validation and testing

3.5.2.1 Details about the Early Stopping algorithm

The early stopping criteria used in this work was defined as:

$$\frac{GL_t}{P_t} > \alpha \quad \text{(Eq. 2.2)}$$

$$GL_t = 100\left(\frac{E_{val,t}}{E_{opt,t}} - 1\right) \quad \text{(Eq. 2.3)}$$

$$E_{opt,t} = \min_{t'<t} E_{val,t'} \quad \text{(Eq. 2.4)}$$

$$P_t = 1000\left(\frac{\sum_{t'=t-k+1}^{t} E_{tr,t'}}{\min_{t'\in(t-k+1;t)} E_{tr,t'}} - 1\right) \quad \text{(Eq. 2.5)}$$

where $E_{tr,t}$ and $E_{val,t}$ are respectively the validation and the training error at iteration t. GL_t is the "generalization loss" at iteration t, defined as the ratio between the current validation error and the minimum validation error since the training started ($E_{opt,t}$). P_t is the "training progress" at iteration t, defined as the ratio between the average and the minimum training error during the last $(t-k)$ iterations (Prechelt, 1998). The GL_t/P_t ratio increases when the generalization loss increases and/or when the decrease in training error gets slower.

3.5.3 Variables and model selection: details and results of cross-validation

K-fold cross-validation (CV) was performed in order to estimate models prediction error. In k-fold CV procedures in general, the dataset is split in k not overlapping subsets, being the "folds", then the model is trained k times, each time considering different (k-1) folds for parameter tuning and the k^{th} fold for validation. The average of mean squared error (MSE) values on the k validation folds is computed as estimate of the average prediction error. The CV process is repeated for different values of the hyperparameters, thus the expected prediction error is estimated at growing model complexity. Typically, the CV error decreases in a first phase, for high values of the shrinking factor or small number of variables or components (low model complexity, underfitting phase). The prediction error starts increasing when the model becomes too complex and overfitting occurs. The minimum CV error corresponds to the optimal hyperparameter value. In this work, due to the small dataset size, leave-one-out CV (LOOCV) was used (Pasini, 2015), in which k is equal to the number of samples. In each iteration of the CV process, only one data was reserved to validation, while the rest was used to tune the model parameters.

Table S2 – ANN training settings and parameters.

Training algorithm	Gradient descent
Learning rate	0.005
Hidden neurons activation function	Hyperbolic tangent
Output layer function	Linear
Max number of epochs (final training after model selection)	500
Early stopping threshold	3

Table S3 – Results about hyperparameters selection after cross-validation of the soft-sensor regressors.

Model	Hyperparameter type	Dataset 1	Dataset 2	Dataset 3
MLR	Number of predictors	3 (TSS, pH, UV abs.)	2 (NO_3^-, NTU)	2 (NTU; UV abs.)
RR	Penalty coefficient	0.0760	0.0946	0.1388
PLS	Number of components	3	4	4
PCR	Number of components	5	5	5
ANN	Penalty coefficients; number of neurons	0.01; 10	0.005; 11	0.0005; 11

Table S4 – Performances of soft sensor models on calibration subsets of dataset 1, 2 and 3.

Model	Dataset 1			Dataset 2			Dataset 3		
	R^2	MAE	MAPE	R^2	MAE	MAPE	R^2	MAE	MAPE
BM1 - base error rate	-	7728	55%	-	193130	290%	-	29894	198%
BM2	-	22133	495%	-	291983	3300%	-	55123	1820%
MLR	0.51	7420	40%	0.39	142420	215%	0.4	22633	103%
RR	0.65	7289	42%	0.22	161827	261%	0.32	23515	177%
PCR	0.64	6980	40%	0.24	152422	206%	0.30	22765	195%
PLS	0.66	6995	41%	0.18	155602	253%	0.33	22659	185%
ANN	0.79	2337	29%	0.71	114515	130%	0.79	11686	47%

Table S5 – Average cross-validation performances of soft sensor models for dataset 1, 2 and 3.

Model	Dataset 1			Dataset 2			Dataset 3		
	R^2	MAE	MAPE	R^2	MAE	MAPE	R^2	MAE	MAPE
BM1 - base error rate	-	7728	55%	-	193130	290%	-	29894	198%
BM2	-	22133	495%	-	291983	3300%	-	55123	1820%
MLR	0.43	7420	40%	0.39	142420	215%	0.4	22633	103%
RR	0.51	7289	42%	0.22	161827	261%	0.32	23515	177%
PCR	0.49	6980	40%	0.24	152422	206%	0.30	22765	195%
PLS	0.53	6995	41%	0.18	155602	253%	0.33	22659	185%
ANN	0.68	2337	29%	0.71	114515	130%	0.79	11686	47%

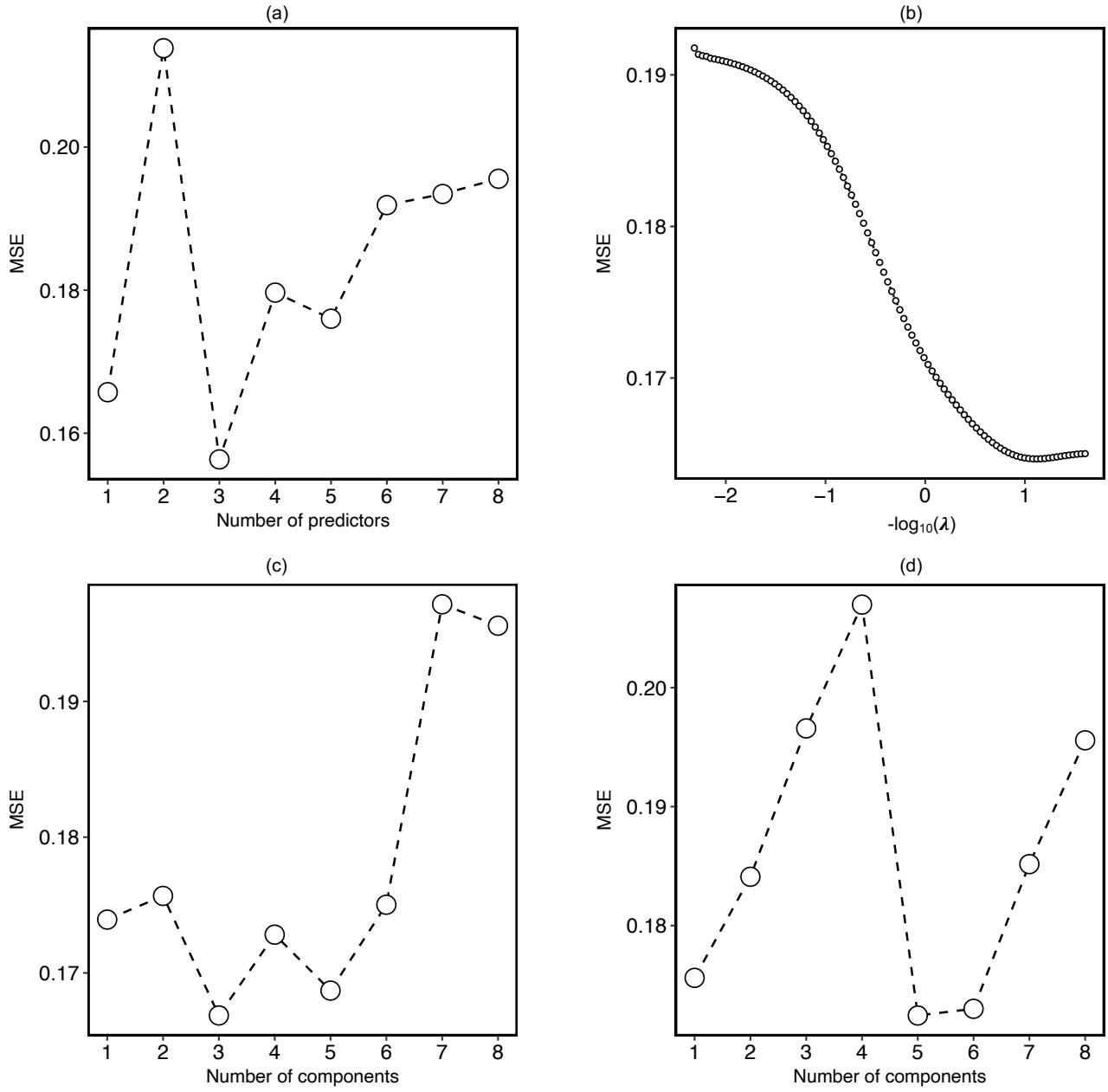

Figure S1 – Results from cross-validation of linear regressors on dataset 1.

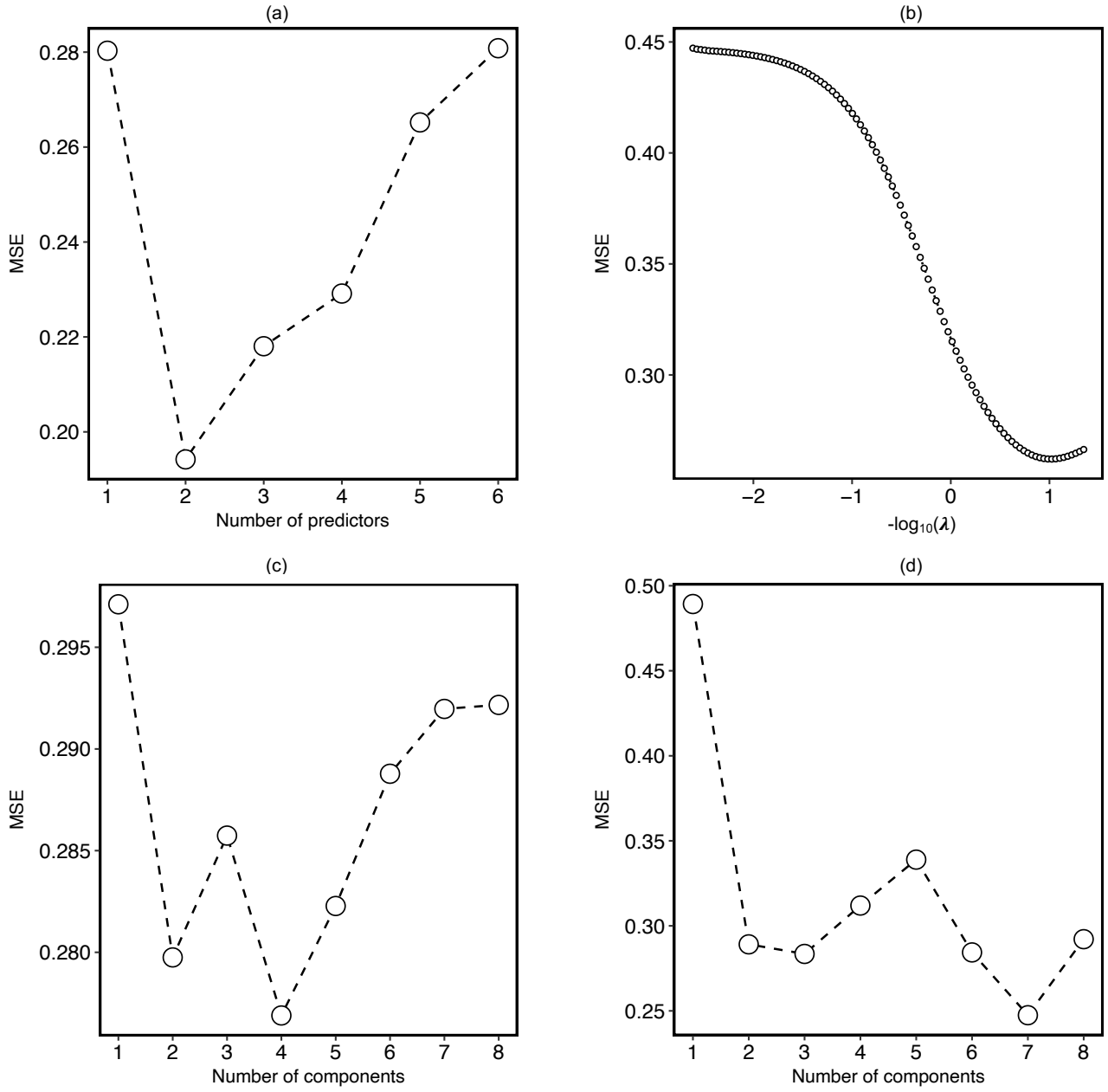

Figure S2 - Results from cross-validation of linear regressors on dataset 2.

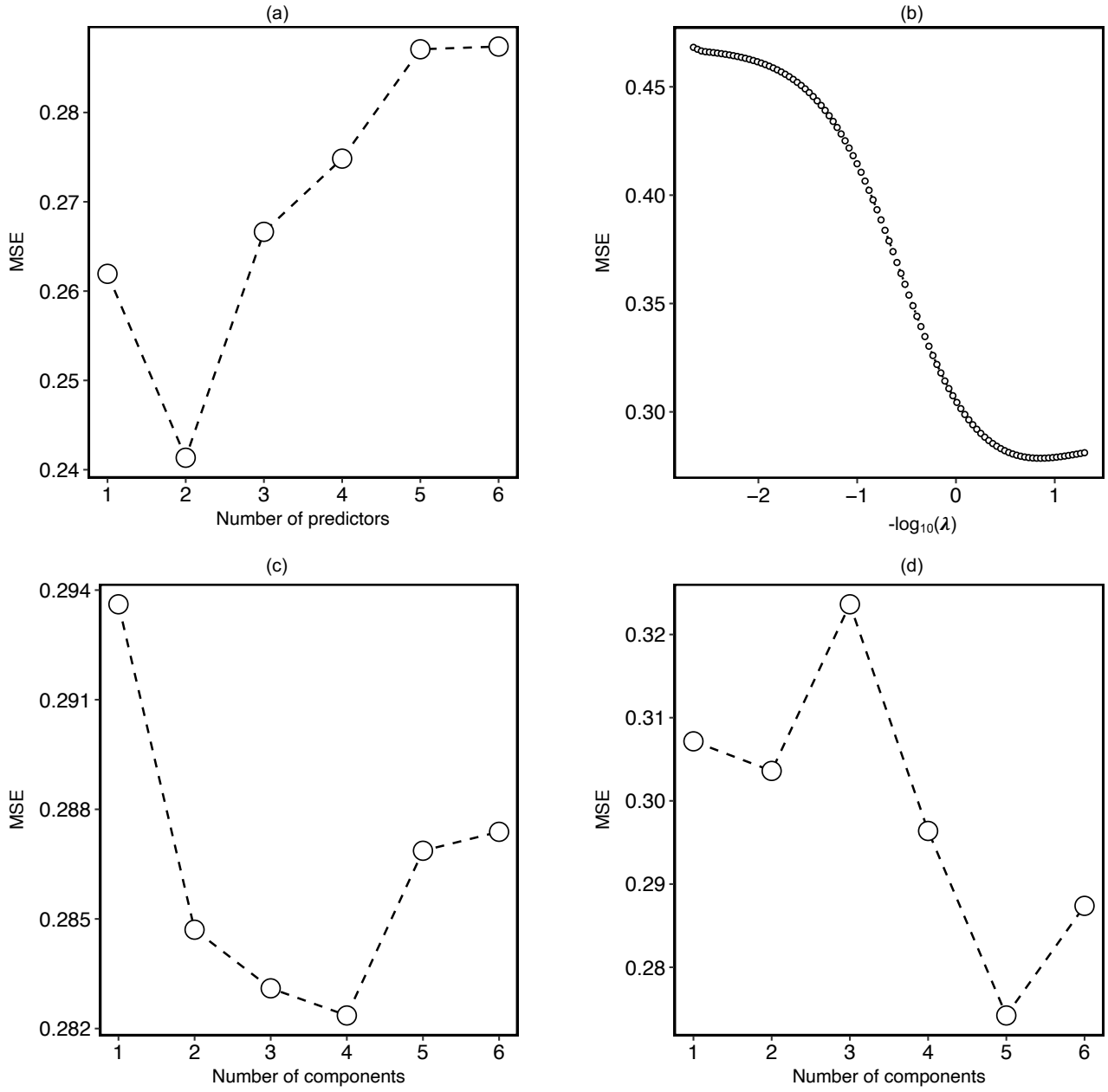

Figure S3 – Results from cross-validation of linear regressors on dataset 3.

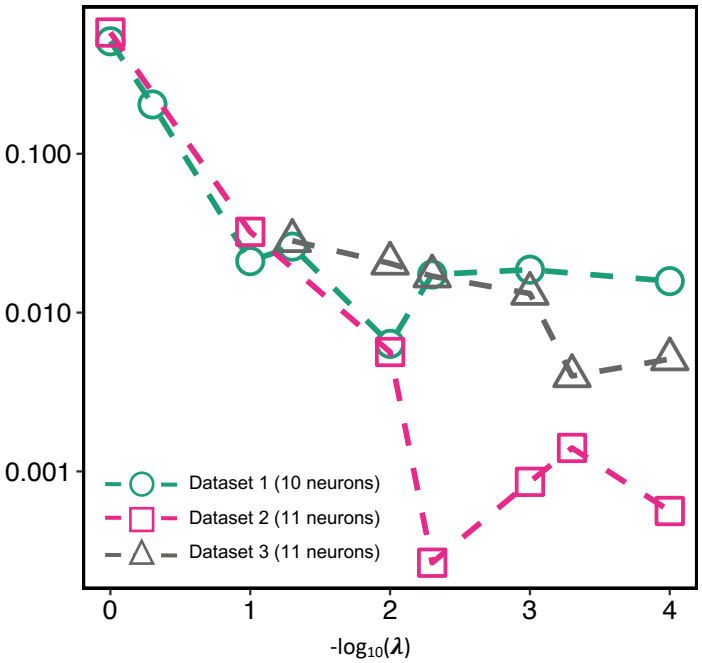

Figure S4 – Results from cross-validation of the ANN models over the 3 datasets. Each point corresponds to the average MSE of the ANN at the given λ and optimized number of neurons.

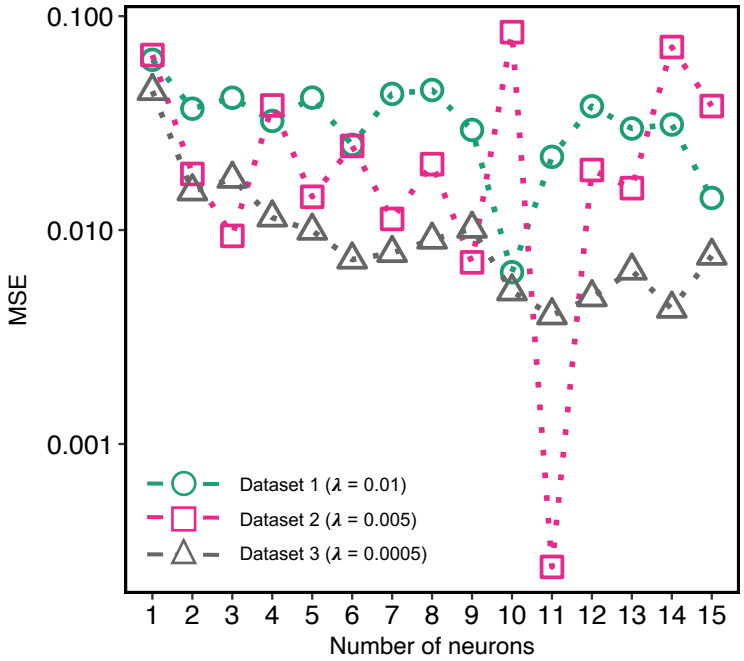

Figure S5 – Results from cross-validation of the ANN models over the three datasets. Each series represents the trend of the average MSE with increasing number of neurons. The three represented series contains the combination of λ and number of neurons, which minimizes the CV error.

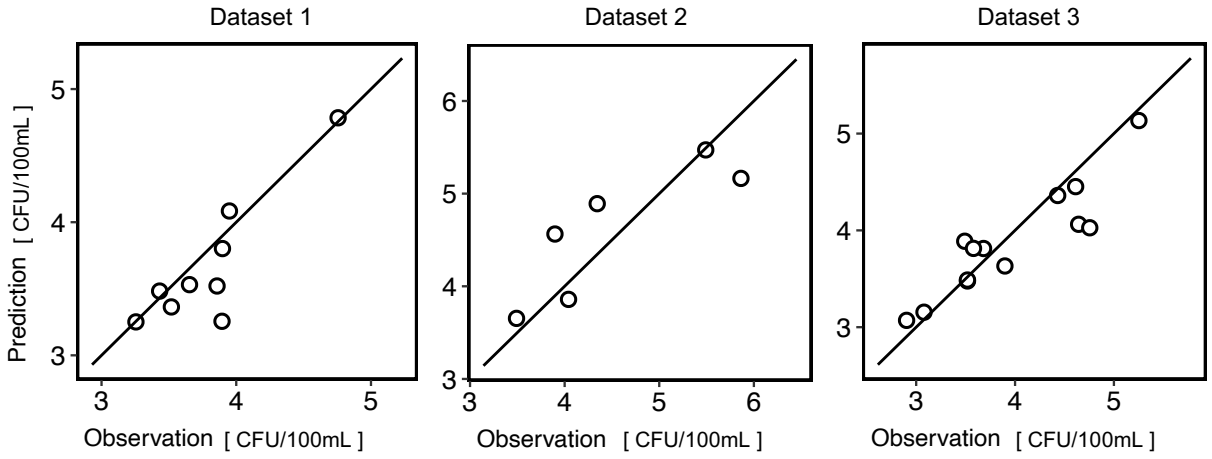

Figure S6 – Test observation vs prediction of the ANN models.

3.5.4 PAA decay experiments: analytical methods and analysis of data

PAA decay batch experiments in the WWTP1 effluent were performed at a PAA dosage of 1 and 2 mg/L. 3 replicates were performed for each PAA dosage. PAA residual concentration was monitored at contact times of 2, 5, 10, 30 and 60 minutes. PAA was measured via DPD colorimetric method (Domínguez-Henao et al., 2018). A first order decay kinetic model was calibrated over observed data via non-linear regression based on gradient descent algorithm. Model equation and parameters are reported in Table S1.

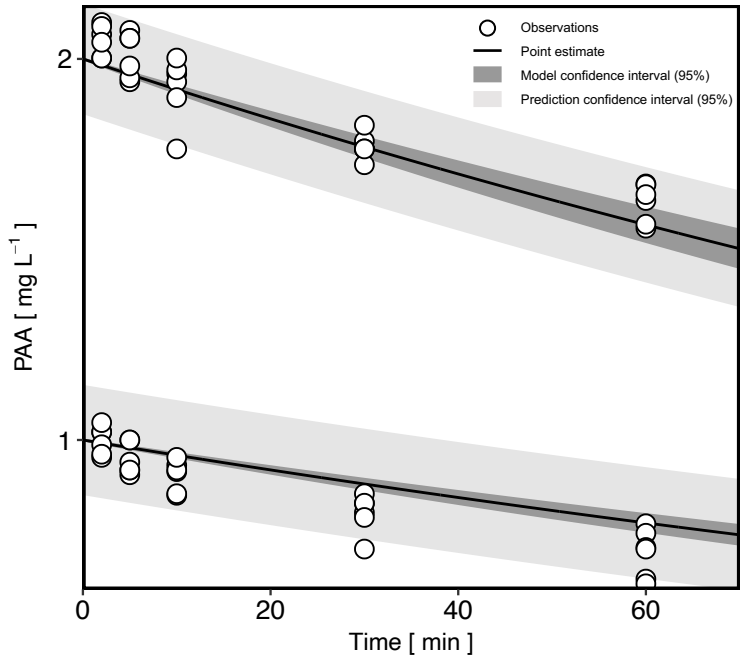

Figure S7 – Results from non-linear regression on batch PAA experiments.

Table S6 – Results from non-linear regression on batch PAA experiments.

	Estimate	95% confidence interval	SE	t statistics	p-value
k	0.0041	0.0036; 0.0045	0.0002	17.516	$4.7353 \cdot 10^{-25}$
Sample size		60			
Model equation		$PAA(t) = PAA_0 e^{-kt}$			

3.5.5 PAA disinfection experiments: analytical methods and analysis of data

PAA inactivation batch experiments of *E.* coli in the WWTP1 effluent were performed at a PAA dosage of 1 and 2 mg/L and contact time between 2 and 60 minutes. Assuming PAA first order decay rate as estimated in PAA decay experiments (see previous paragraph), effluent samples were exposed to a disinfectant dose up to about 120 mg L^{-1} min. Inactivation model developed by Domínguez Henao et al. (2018) was calibrated over observed data via non-linear regression based on gradient descent algorithm. Model equation and parameters are reported in Table S2.

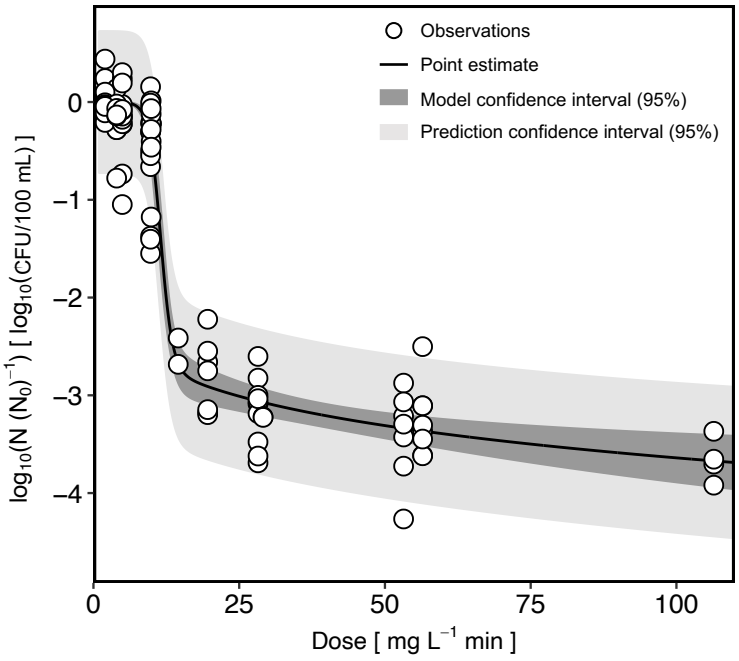

Figure S8 - Results from non-linear regression on batch disinfection experiments.

Table S7 - Results from non-linear regression on batch disinfection experiments.

	Estimate	95% confidence interval	SE	t statistics	p-value
k'	4.4561	3.3703; 5.5420	0.54726	8.1427	$1.1862 \cdot 10^{-12}$
n	0.1372	0.0724; 0.2020	0.03265	4.2007	$5.8283 \cdot 10^{-05}$
h	11.476	11.0422; 1.9098	0.21862	52.492	$4.4266 \cdot 10^{-74}$
Sample size		102			
Model equation		$\frac{\ln(N)}{\ln(N_0)} = \frac{-k' D^n}{1 - e^{h-D}}$			

3.5.6 Additional details on collected data

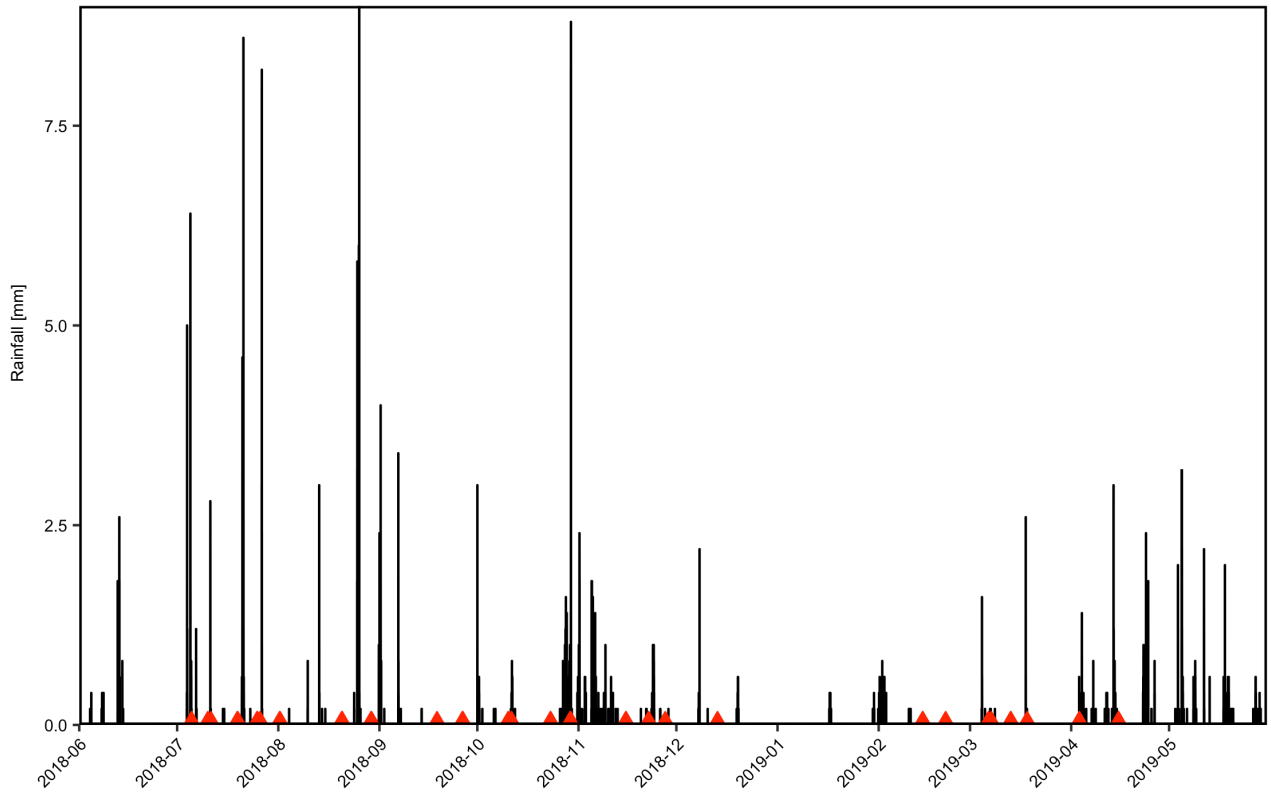

Figure S9 – Rainfall data during collection campaign in WWTP1. Red marks correspond to sampling times.

3.5.7 Additional notes on the application of sensitivity analysis

This family of methods is based on the assumption that the more important is one input in determining the model output, the higher is the reduction in output variance when that input is fixed to its true value. The main advantage of this approach is that the analysis accounts for both the effect of the marginal influence of each single predictor, as well as the effect of their joint interactions. Moreover, the results validity is extended to all the variability space of the predictors. Another advantage of VSA is that it works regardless the model properties (Saltelli et al., 2006).

References

Domínguez-Henao, L., Turolla, A., Monticelli, D., & Antonelli, M. (2018). Assessment of a colorimetric method for the measurement of low concentrations of peracetic acid and hydrogen peroxide in water. Talanta, 183(February), 209–215. https://doi.org/10.1016/j.talanta.2018.02.078

Domínguez Henao, L., Cascio, M., Turolla, A., & Antonelli, M. (2018). Effect of suspended solids on peracetic acid decay and bacterial inactivation kinetics: Experimental assessment and

definition of predictive models. Science of the Total Environment, 643, 936–945. https://doi.org/10.1016/j.scitotenv.2018.06.219

Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. Technometrics, 35(2), 109–135. https://doi.org/10.1080/00401706.1993.10485033

Hastie, T. et. all. (2009). Springer Series in Statistics The Elements of Statistical Learning. The Mathematical Intelligencer, 27(2), 83–85. https://doi.org/10.1007/b94608

Pasini, A. (2015). Artificial neural networks for small dataset analysis. Journal of Thoracic Disease, 7(5), 953–960. https://doi.org/10.3978/j.issn.2072-1439.2015.04.61

Prechelt, L. (1998). Automatic early stopping using cross validation: Quantifying the criteria. Neural Networks, 11(4), 761–767. https://doi.org/10.1016/S0893-6080(98)00010-0

Saltelli, A., Ratto, M., Tarantola, S., & Campolongo, F. (2006). Sensitivity analysis practices: Strategies for model-based inference. Reliability Engineering and System Safety, 91(10–11), 1109–1125. https://doi.org/10.1016/j.ress.2005.11.014

Zhou, D. X. (2013). On grouping effect of elastic net. Statistics and Probability Letters, 83(9), 2108–2112. https://doi.org/10.1016/j.spl.2013.05.014

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 67(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Chapter 4: Artificial neural network modeling of full-scale UV disinfection for process control aimed at wastewater reuse

Abstract

Accurate modeling of wastewater ultraviolet disinfection is fundamental as support for process optimization and control. Detailed modeling of hydrodynamics and fluence rate via computational fluid dynamics, coupled to laboratory studies of inactivation kinetics, are usually the preferred approach for UV disinfection modeling. Despite this approach often provides accurate predictive performance, it requires significantly high computational time, making it unfeasible for real-time process control. In this study, to enable an effective process control, black-box regression models were assessed as a modeling alternative for UV disinfection, synthesizing hydrodynamics, fluence rate and inactivation kinetics. UV disinfection of a full-scale wastewater treatment plant in Italy was monitored for 10 months, measuring influent and effluent *E. coli* concentration, turbidity, absorbance at 254 nm, temperature and flow rate at different UV doses. Considering the usually observed distribution of effluent *E. coli* concentration and the zero inflation of the collected dataset, Poisson, zero-inflated Poisson and Hurdle generalized linear models were tested, as well as two-part models coupling a classifier describing the *E. coli* zero-count events and a regressor estimating the magnitude of *E. coli* concentrations in positive-count events. The two-part artificial neural network model showed the best predictive performance, being able of both describing nonlinearities and handling the high proportion of null values in the dataset. The deployment of this model to control ultraviolet disinfection was simulated, estimating a plausible 63% energy saving.

Keywords: UV disinfection; artificial neural network; control; *E. coli*; wastewater; reuse; zero inflated dataset.

This chapter has been published on “Journal of Environmental Management”.

4.1 Introduction

Wastewater ultraviolet (UV) disinfection provides high inactivation levels for many microbial indicators (Lazarova et al., 1999), at the same time avoiding the drawback of by-products formation,

typical of chemical disinfection, which is the most widespread disinfection alternative. However, UV disinfection optimization requires to address the trade-off between the maximization of inactivation efficiency and process reliability and the minimization of energy consumption.

Meeting the selected inactivation target requires the optimal adjustment of UV radiation intensity, which is the only quantity which can be controlled in a UV disinfection system, but only by the support of a proper process model, accounting for reactor hydrodynamics, UV fluence rate and inactivation kinetics. The description of these three elements introduces several modeling challenges. As for UV reactor hydrodynamics, it is affected by flow rate and configuration of contactor and lamps; two different approaches can be adopted to predict the exposure time to UV radiation of fluid particles, differing in complexity, description ability and computational needs: i) computational fluid dynamics (CFD), which is a time and computationally intensive approach (Wols et al., 2012), ii) conceptual hydraulic modeling, based on a combination of ideal reactors (Fenner et al., 2005), which is a simpler alternative in terms of model complexity. Anyway, both these approaches require to perform *ad hoc* tracer experiments. Once fluid dynamics is known, the second step is the description of the radiation transfer in the wastewater matrix, which determines the fluence rate. The most detailed approach to describe UV fluence rate is again CFD modeling (Chen et al., 2011). A simpler alternative is modeling UV radiation as the result of the contributions of multiple point or segment sources (Powell et al., 2015); however, this latter approach, independently of the adopted modeling method, neglects the impact of fluid dynamics. It should be stressed that wastewater quality critically affects the actual UV radiation intensity distribution in the reactor volume. In detail, wastewater absorbance at UV wavelengths increases due to many dissolved compounds, particularly due to organic matter (Bolyard et al., 2019), thus reducing effective UV radiation passing through water and irradiating microbes. Moreover, suspended solids reduce actual UV dose by scattering, reflecting and refracting light (Jolis et al., 2001; Madge et al., 2006; Carrè et al., 2018). Finally, the third step pertains to inactivation kinetics, which are highly nonlinear and affected by significant uncertainty, due both to the intrinsic variability of microbial concentration and reaction with the disinfectant, and to the shielding effect of suspended solids (Chahal et al., 2016). Site specific inactivation kinetics are usually determined by collecting wastewater samples and performing batch disinfection experiments, in which they are exposed to UV radiation of defined intensity for a known time through a collimated beam apparatus (Bolton and Linden, 2003).

Recently, many of these modeling and experimental techniques were integrated to build up modeling frameworks to predict UV disinfection efficiency, while accounting for the uncertainty due to modeling and wastewater quality variability (Ahmed et al., 2019). Since most of these

modeling approaches are based on CFD, both for hydrodynamics and fluence rate description, they imply high computational time and effort. Running a CFD simulation of a UV reactor, given fixed operating conditions and geometric configurations, can take hours or days. Thus, while CFD-based models work for static modeling of UV disinfection efficiency, even within a probabilistic approach, they are not suitable for process control applications, where the disinfection efficiency has to be iteratively estimated every time new information on wastewater flow rate and quality is acquired. In fact, since the average residence time of UV reactors is usually in the order of seconds or minutes, a significant time lag between model predictions based on CFD simulations and effective plant output is inevitable. An affordable solution is represented by a black-box approach directly modeling the relationship between UV radiation intensity, flow rate, wastewater quality characteristics and inactivation efficiency, providing a computationally light model suitable for real-time control deployment. This solution only requires the collection of enough data in different operating conditions of the disinfection unit and to perform a regression analysis. In recent years, some studies highlighted the potential of this simpler but effective approach, but case studies are limited to the lab and pilot scale (Lin et al., 2012) or to "surrogate" modeling approaches, where statistical models were trained over mechanistic model simulations, in order to mimic them while reducing computation time (Xu et al., 2015).

This study proposes a black-box approach for modeling a full-scale UV disinfection for *E. coli* inactivation in wastewater, predicting the UV system efficiency by considering number of active UV lamp banks, flow rate, initial *E. coli* concentration, UV absorbance, turbidity and temperature as predictors. Linear and nonlinear regressors were tested to capture input-output relationships from collected data, synthesizing the effect of hydrodynamics, wastewater quality and inactivation kinetics in a single black-box mathematical model. Special attention was given to mathematical treatment of zero-count data of effluent *E. coli* concentration, which are typical of high efficiency disinfection systems. The developed approach was validated in a full-scale UV disinfection unit in S. Rocco WWTP (Milan), one of the biggest in Italy (1.050.000 PE), by measuring *E. coli* concentrations at different operating conditions in terms of UV radiation, flow rate and wastewater quality. The accuracy of the best regression model was tested on the data collected by the WWTP utility during an irrigation season. Moreover, within the same time window, real-time control of UV disinfection based on the black-box regression model was simulated, considering *E. coli* discharge limit imposed by Italian law for agricultural reuse. Potential energy savings enabled by the control strategy were also estimated.

4.2 Materials and Methods

4.2.1 Case study and data collection

Samples were collected between June 2018 and April 2019 from a S. Rocco WWTP (1.050.000 PE), in the Milan urban area. The WWTP is based on activated sludge process for nitrification and denitrification, followed by sand filtration and UV disinfection. The UV disinfection stage is made of two parallel lines. This study focused on one line, which is made of three identical parallel open channels, each one equipped with three banks containing 144 monochromatic (254 nm) low pressure/high intensity UV lamps each (378 lamps per line, 1134 in total), for a total of about 136 kW of installed power per channel. In each channel, the UV fluence rate can be adjusted by changing the number of active UV lamp banks. Thus, each channel can work at three discrete levels of nominal UV dose. The disinfection unit was designed to deliver 19, 38 and 57 mJ cm^{-2} when operating with one, two and three active lamp banks per channel, at the nominal flow rate of 4 m^3 s^{-1}.

The dataset used for model calibration was collected in 129 full-scale disinfection experiments. In each experiment, wastewater at the inlet and the outlet of one of the disinfection channels was sampled and analyzed in laboratory for *E. coli* concentration and absorbance at 254 nm (UV254). Samples were collected under different operating conditions, meaning with one, two and three active UV lamp banks. Turbidity, temperature and flow rate data were collected at a 10-minutes frequency during all the campaigns by on-line sensors.

A second dataset, used for model testing, was collected during the routine UV disinfection monitoring carried out by WWTP managers at a roughly 3-days frequency, during the 2018 irrigation season (from June to August). Flow rate, temperature, turbidity and residual *E. coli* at the UV disinfection outlet were monitored during this time window.

4.2.1.1 Sampling procedure, physical, chemical and microbiological analysis

All samples were manually collected in 1-L sterile amber bottles, transported to the laboratory in refrigerated bags (within 30 minutes from sample collection) and processed immediately at the WWTP laboratory. *E. coli* were enumerated by membrane filtration method according to Standard Methods (APHA/AWWA/WEF, 2012), using 0.45 μm pore size cellulose nitrate membranes (Whatman) and chromogenic agar (Microinstant® Chromogenic Coliforms Agar, Scharlau) as culture medium. Inoculated plates were incubated at 37 °C for 18-24 h. Turbidity, temperature and flow rate data were monitored by on-line sensors (micro::station®, S:CAN). UV254 (1-cm optical

path) was measured by DR6000 laboratory spectrophotometer (HACH) on samples filtered by 0.45 μm pore size cellulose nitrate membranes (Whatman).

4.2.2 UV disinfection regression model development

4.2.2.1 Data preprocessing

Input data were standardized to have 0 mean and standard deviation equal to 1, in order to ensure they have equal importance in the calibration process.

4.2.2.2 Regression models identification

Five linear regression modeling techniques were evaluated to estimate *E. coli* concentration in UV disinfected effluent, being: (i) multiple linear regression (MLR) via ordinary least squares (OLS); (ii) generalized linear model (GLM) based on Poisson distribution (GLMP); (iii) GLM based on zero-inflated Poisson distribution (GLMZIP); (iv) Hurdle model based on Poisson distribution (HMP); (v) a two-part model (TPM) made of a sequence of a logistic binary classifier (LBC) and a MLR.

GLMP assumes a Poisson conditional distribution of the response, according to a model equation called "exponential mean function" (Lambert et al., 1992):

$$f_P(y|x) = \frac{e^{-\mu}\mu^y}{y}, \tag{2.1}$$

$$E[y|x] = \mu = e^{x\beta} \tag{2.2}$$

where y is the response, x is the vector of the predictor, β is the vector of model parameters, f_P indicates the Poisson probability mass function and μ is the Poisson distribution parameter. Equation 2.2 defines the conditional mean of the response, given the predictors.

GLMZIP[16] provides a way to model count data with excess zeros. In fact, assuming a Poisson distribution, a separate component is added to inflate the probability of zero-count events. Zero-count events arise then from both a point mass concentrated on zero and the Poisson distribution, as described by the expression of the zero-inflated Poisson probability mass function (f_{ZIP}):

$$f_{ZIP}(y|x) = \begin{cases} \pi + (1-\pi)f_P(y) & y = 0 \\ (1-\pi)f_P(y) & y > 0 \end{cases} \tag{2.3}$$

where the probability π of having zero-count event depends on predictors as follows:

$$ln\frac{\pi}{1-\pi} = x\beta \tag{2.4}$$

The conditional mean of the response is again defined by equation 2.2.

HMP (Cragg et al., 1971) is a sequential model where a binomial probability governs the binary outcome of whether a count has a null or a positive realization. If the realization is positive, its

conditional distribution is governed by a truncated at zero Poisson distribution. The HMP model can be written as:

$$f_H(y|x) = \begin{cases} \pi & y = 0 \\ (1-\pi) f_{TP}(y) & y > 0 \end{cases} \quad (2.5)$$

$$ln\frac{\pi}{1-\pi} = x\beta \quad (2.6)$$

$$E[y|x] = \mu = e^{x\beta} \quad (2.7)$$

where $f_{TP}(y)$ is the zero-truncated Poisson distribution.

The TPM is an alternative sequential model, which couples an LBC and an MLR model. The LBC was calibrated on the whole dataset to tell between event with null and positive concentration of *E. coli*. A MLR model was then calibrated only on data corresponding to positive *E. coli* effluent concentrations, as an estimator of the concentration magnitude. For MLR calibration $\log_{10}$ transformation of *E. coli* concentrations was used, in order to equally weight small and large values (Shu and Burn, 2004). When the model is used to predict, given a set of predictor observations, the LBC estimates if the concentration of *E. coli* is null or positive, while, in case of positive concentration events, the MLR model predicts the magnitude of concentration.

Beyond these five linear modeling techniques, two nonlinear options were tested, being an artificial neural network (ANN) model and a nonlinear two-part model (TPANN) consisting in the sequence of two ANNs, working respectively as a classifier and a regressor, in the same manner as the TPM. ANN was chosen as nonlinear family of models since it can reproduce any nonlinear relationship at the price of increasing its complexity (Hornik et al., 1990), i.e. increasing the number of layers and neurons per hidden layer. In this work, the "shallow" ANN model was chosen, being a network with a three-layer architecture, made by one input, one hidden and one output layer.

All the seven models were calibrated and validated according to a 10-folds cross-validation (CV) approach (Hastie et al., 2009). To perform the CV, the dataset was split in 10-not-overlapping subsets, being the "folds", then the model was trained 10 times, each time considering 9 different folds for parameter tuning and the 10th fold for validation. The average performance values on the 10 validation folds were computed as estimate of the average prediction error. Mean Absolute Percentage Error (MAPE) was computed as performance metric on validation folds. MAPE was chosen for its easy interpretability and to balance the importance of large and small values in determining the average error.

In the case of ANN and TPANN, CV was also used to optimize the number of neurons. The CV process was repeated for increasing number of neurons, from 1 to 25, thus the expected prediction error was estimated at growing model complexity. Typically, the CV error decreases in a first phase, for small numbers of neurons (low model complexity, underfitting phase). The prediction

error starts increasing when the model becomes too complex and overfitting occurs. The minimum CV error corresponds to the optimal number of neurons. Even in single hidden layer network, the number of weight parameters defining the model can be very high, being equal to $p(n+1)$, where n is the number of hidden neurons and p is the number of inputs. In small datasets, such as in this work, the number of weights of a shallow ANN with a few hidden neurons can easily become comparable to the number of observations and predictors, and even outnumber them. The consequence is to produce a model which easily overfits calibration data and provides very poor generalization capability. For this reason, in this work the early stopping technique was used when training ANNs, to avoid overfitting and improve generalization. Validation error was monitored during each training process. Training and validation errors usually decrease together during the first training iterations; at later stages of training, the network starts to overfit the data and the validation errors typically rise. The training process is stopped once the validation error rises for more than 10 subsequent iterations. The values of parameters at the minimum validation error are kept as the result of training. Optimization of the number of neurons was repeated for three different kinds of activation functions, being sigmoidal, hyperbolic tangent and radial basis functions.

In the ANN training process, the Levenberg-Marquardt backpropagation algorithm was used for MSE minimization.

Linear models were developed in R 4.0.1 environment, using package `pscl 1.5.5` for zero-inflated Poisson regression and hurdle model regression. ANN models were developed in MATLAB R2018b, using the Deep Learning Toolbox.

4.2.2.3 Model testing and simulation of UV disinfection control

The model with the best CV performance was tested on effluent *E. coli* concentration data collected between June and August 2018. During this period the WWTP delivered reclaimed wastewater for reuse in the agricultural district in the southern Milan area. Since the Italian discharge limit for agricultural reuse is 10 CFU/100 mL, the disinfection line under study operated with all the UV lamp banks switched on. *E. coli* concentrations at the inlet of the UV disinfection line were estimated each 10 minutes by using a model from a previous study (Foschi et al., 2021). Turbidity, flow rate and temperature were available from WWTP sensors records. UV absorbance was assumed equal to the mean observed during data collections, since it was not real-time monitored by the WWTP.

95% confidence bounds of the model estimate (*E. coli* concentration at the UV disinfection unit outlet) were generated via “bootstrap aggregation”, or “bagging”, approach (Breiman, 1996). It consists in randomly resampling with replacement new datasets from calibration data, each sample

having the same size as the original calibration dataset. In each bootstrap sample, a different model was calibrated and the probability distribution of model parameters was estimated by the observed distribution of parameters coming from each model training. Similarly, both a point estimate, by averaging predictions, and a confidence interval, by taking the observed percentiles of predictions, can be estimated by predictions from all bootstrapped samples (Efron & Hastie, 2016).

The control of the UV disinfection process based on the regression model was then simulated. The UV model was used to estimate at each time the minimum number of UV lamp banks to be activated to guarantee the desired probability of non-compliance (PONC) (Talebizadeh et al., 2014) with the 10 CFU/100 mL limit.

4.3 Results and Discussion

4.3.1 Preliminary data analysis

Boxplots of *E. coli* concentration before and after UV disinfection are reported in Figure 1 and descriptive statistics are summarized in Table 1, considering the effect of the different number of active UV lamp banks. The effect of the increase of UV radiation is straightforward by looking at the trend of *E. coli* average values. Most of inactivation is reached with just one active UV lamp bank, maintaining most of the time *E. coli* concentration below 100 CFU/100 mL. However, activating the second and the third UV lamp banks seems fundamental to safely maintain concentrations below the 10 CFU/100 mL agricultural reuse limit in all cases. It is evident by looking at Figure 1 and Table 1 how, given a number of active UV lamp banks, data are dispersed around the mean. The number of operating UV lamp banks is then not sufficient to measure the degree of inactivation with proper accuracy, since they do not univocally define the UV dose which wastewater is exposed to. In fact, flow rate values are needed to derive exposure time and information about wastewater quality are fundamental to measure the amount of UV radiation which is absorbed and reflected by wastewater, making it not available for bacterial inactivation.

Table 1 – Summary statistics about measured *E. coli* effluent concentration (CFU/100mL) in the wastewater flowing through one UV channel as a function of the number of operating UV lamp banks. Influent *E. coli* concentration corresponds to zero operating UV lamp banks. Sample size: 43 data for each operating condition.

Number of operating banks	Min.	Avg.	Max.	Std. Dev.	Coeff. of Var.	Proportion of zero-count data
0	600	7,182	57,000	10,428	1.4	0%
1	0	45.2	810	147.6	3.3	12%
2	0	24.1	510	104.5	4.3	42%
3	0	0	8	2	2.5	60%

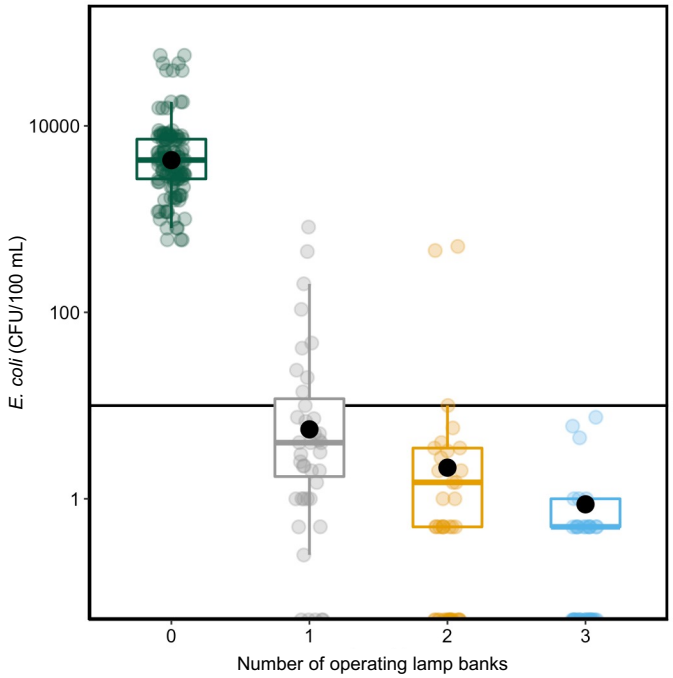

Figure 1 – Measured *E. coli* effluent concentration as a function of the number of operating UV lamp banks. Values corresponding to zero operating UV lamp banks are influent *E. coli* concentrations. The agricultural reuse discharge limit of 10 CFU/100 mL is marked as a black solid line.

These considerations are supported by the observed trends of effluent *E. coli* concentration data with flow rate, temperature, turbidity, UV254 and initial *E. coli* concentration, reported in Figure 2. By looking at regression lines slopes, the effects of these factors are always present with one operating UV lamp bank, with the exception of UV254, seeming relevant even with 2 and 3 operating UV lamp banks. Having in mind a typical UV inactivation kinetic curve, composed of a linear and a tailing-off segment, a probable reason is that a single operating bank provides a UV dose which roughly corresponds to the linear segment, where the inactivation rate is higher (*inter alia*: Antonelli et al., 2008, Carré et al., 2018; Ahmed et al., 2019). Thus, changes in residence time (which varied between 7 and 46 s due to changes in flow rate) and in wastewater characteristics impact on the level of inactivation by moving the applied UV dose along the linear segment, resulting in significantly different output *E. coli* concentrations. On the contrary, UV dose is located in the tailing-off segment with three active UV lamp banks, and its change does not affect significantly the inactivation rate. As expected, an increase in turbidity, UV254 and flow rate determines a decrease of the inactivation efficiency. Temperature causes the opposite, although it is known how its impact on UV disinfection is usually negligible (Severin et al., 1983; Abu-ghararah et al., 1994). In this case, temperature could have a role in explaining disinfection efficiency as a proxy variable of upstream processes working. As first guess, lower temperatures could determine lower activated sludge performance and higher concentration of soluble organics in the disinfection influent. This hypothesis is supported by the observed correlation between temperature and

ABS254 data (-0.54, p-value = $1.2 \cdot 10^{-10}$). Differently, most of the time, when more than one UV lamp bank is active, flow rate, turbidity and temperature seem not to affect inactivation. Finally, it was expected that the influent *E. coli* concentration is positively correlated to the effluent *E. coli* concentration.

Overall, it is clear how none of the considered factors can exhaustively describe alone the UV disinfection process, which is conversely the result of the sum and interactions of all these factors, suggesting that a multivariate approach is the most appropriate to describe it.

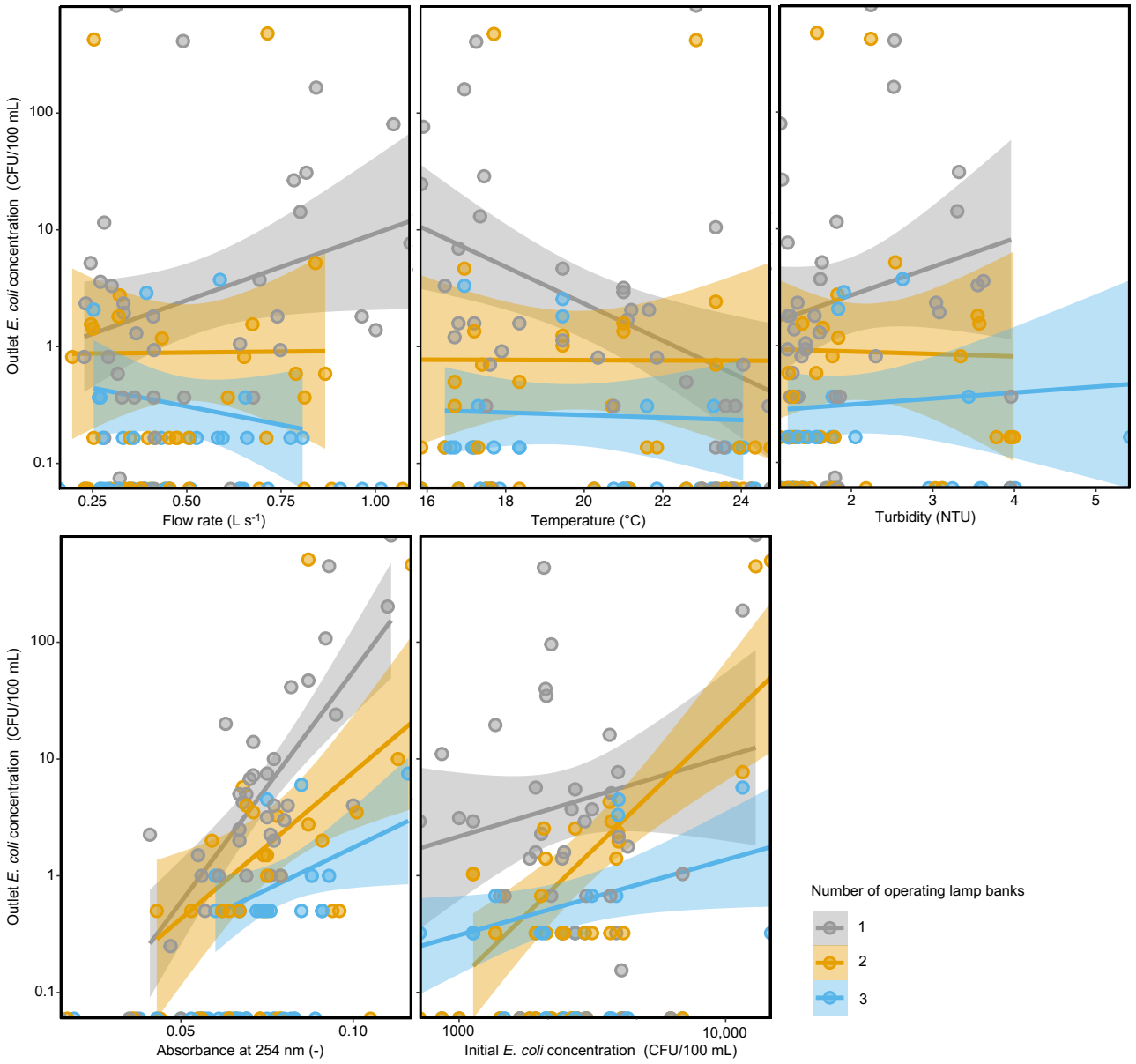

Figure 2 –Outlet *E. coli* concentration after UV disinfection vs. operating conditions (flow rate) and wastewater characteristics, as a function of the number of operating UV lamp banks. Linear trends are obtained by linear regression (shaded areas represent 95% confidence intervals).

4.3.2 Model selection and testing

In the model identification step, several linear and nonlinear regression models were calibrated and their prediction performances were compared in order to find the best predictive tool to support the control of the disinfection line. All the monitored factors, i.e. number of active UV lamp banks, flow rate, influent *E. coli* concentration, turbidity, UV254 and temperature, were selected as model predictors, since, as reported in paragraph 3.1, they all showed relationships with the model output (effluent *E. coli* concentration) which were also supported by *a priori* knowledge about UV disinfection (Carrè et al., 2018; Bolyard et al., 2019). Besides being the most important factor influencing UV disinfection efficiency, all these parameters, except for influent *E. coli* concentration, are easy to be monitored on-line for real-time control applications. The choice of the modeling approaches to be compared was influenced by the presence of a high number of counts equal to zero in effluent *E. coli* concentrations, consisting in about 38% of the total. The number of zero-count events rises with the number of active banks, as reported in Table 1. In order to include zero-count data in the analysis, not to lose the information related to when zero-count events occur, several regression approaches specific for zero-inflated datasets were selected and compared. The comparison of the regression techniques by the CV performance of the studied models is reported in Figure 3. The MLR model was included in the analysis as a benchmark, being the simplest and most widespread multivariate linear regression tool. Due to the leverage effect of the large amount of zeroes and of the few extreme values (see boxplots in Figure 1), the MLR model performed poorly and is then not adequate for prediction purposes. The GLMP leads to better performance, since it assumes the conditional distribution of the output to be a Poisson, which is defined for non-negative integer values, allowing for zero-count events, and it can describe right skewed observed distributions. Adding a point mass concentrated on zero by a zero-inflated Poisson model does not improve prediction performance. The use of HMP leads to an additional reduction of error. HMP model probably improves performance with respect to ZIP, since it considers the processes generating null and positive counts as separated (Colin et al., 2013). Probably for the same reason, the TPM leads to even better performance, the best among studied linear models. A possible additional reason of this improvement is the fact that the linear regression sub-model of the TPM, estimating the magnitude of positive concentrations, was calibrated only on the strictly positive observations of the dataset; thus, logarithmic transformation of *E. coli* concentrations was possible, which better balanced the relative importance of observations in the calibration phase.

The ANN model shows better performance than the MLR, indicating that nonlinear input-output relationships and interactions among predictors are important in predicting UV efficiency.

However, despite it is a much more complex models, ANN performance is similar to TPM, since it does not treat separately the occurrence of zero-count events. Actually, the best performance among all tested models is displayed by the TPANN model, which dedicates a sub-model to the classification of null and positive count events and properly identifies nonlinearities in data.

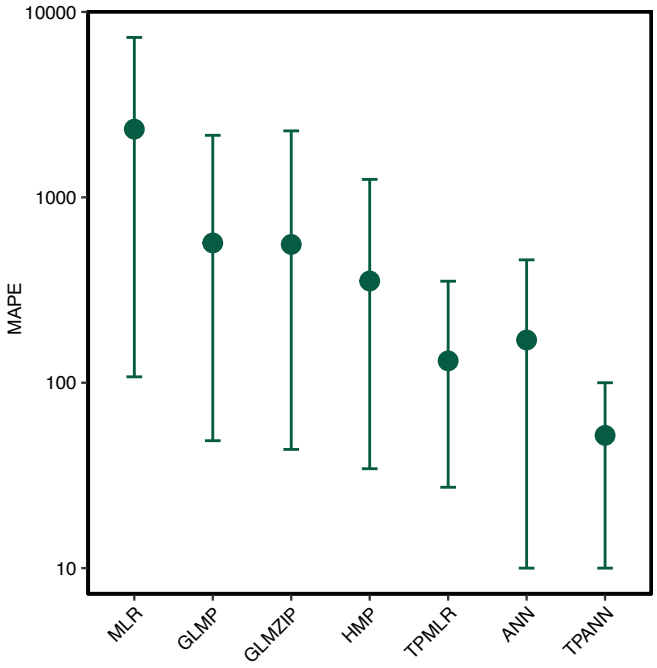

Figure 3 – Predictive performance, expressed as MAPE, from CV of the different studied regression models.

ANN and TPANN predictions on all CV validation folds are summarized and compared in Figure 4.a. The two models perform similarly on large concentration values, with slightly better performance of the ANN model. Differently, the TPANN model is significantly more accurate on middle and low concentration values. Moreover, as reported in Figure 4.b, the TPANN model is very accurate in predicting null concentration events (accuracy of 95%), while ANN predictions in those cases are very dispersed and, inevitably, negative about half of the times. Better performances of the TPANN model are then due to: (i) the use of ANN models, capable of capturing nonlinear relationships in data; (ii) the two-part structure, which allowed to dedicate a sub-model to the identification of null concentration events and to apply a logarithmic transformation of positive counts data, similarly to the TPM case described before.

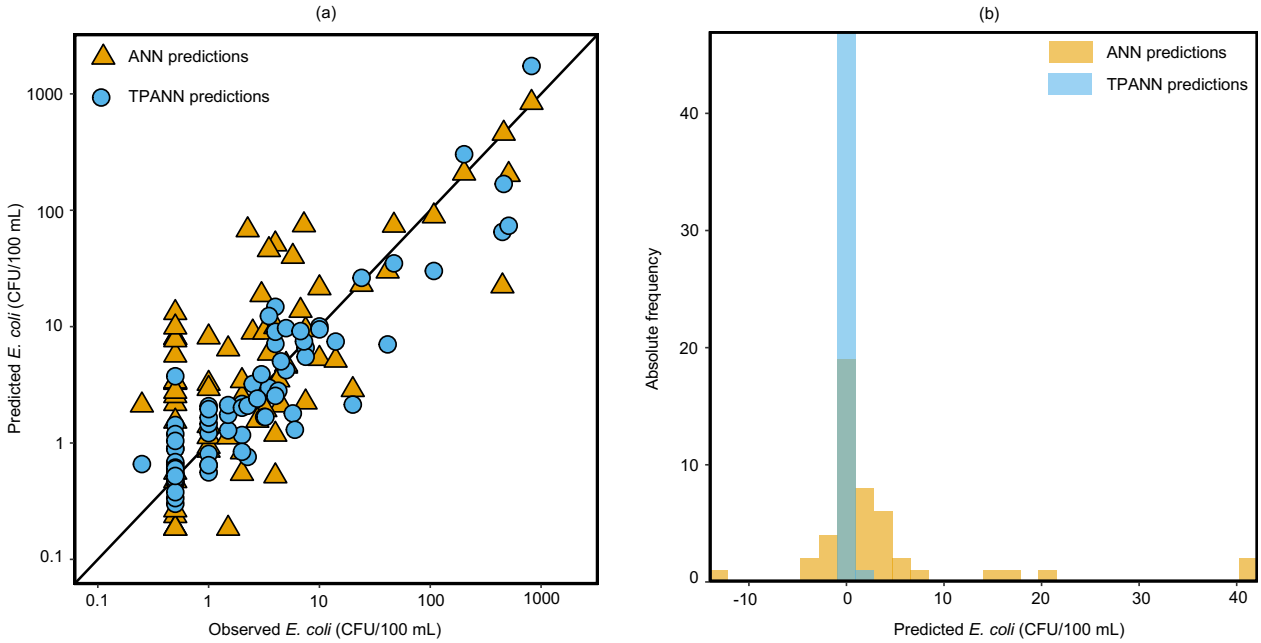

Figure 4 – Comparison of the predictions of the ANN and TPANN models, when observed *E. coli* effluent concentration is: (a) ≥0; (b) = 0.

The TPANN model is then definitely the most suitable for modeling and control purposes in the developing field of wastewater reuse, where the collection of zero-inflated datasets is a straightforward consequence of the goal these processes are designed for. This black-box approach provides an effective predictive tool avoiding the complex and computationally intensive explicit modeling of hydrodynamics and UV fluence rate, saving time and effort during model identification and unlocking process control applications, which requires low computation time. In fact, while running a CFD model can take hours, getting a prediction with such kind of statistical model is a matter of seconds. The calibration of the regression models needs to collect data about the UV process operations, which could even come from the routine monitoring of the unit by the WWTP operators and from available on-line sensors. This data collection effort replaces more expensive and demanding *ad hoc* experiments like multiple collimated beam experiments performed at different wastewater qualities, tracer studies and/or biodosimetry tests.

Predictions of the TPANN model were tested over data which were "unseen" by the model, between June and August 2018, as reported in Figure 5. Measured effluent *E. coli* concentrations revealed that the agricultural reuse limit was never exceeded, but concentrations increased and approached the limit a few times between July 15th and 20th. In fact, in this period the on-line sensor registered an increase in turbidity levels, probably due to the occurrence of some rain events, which affected the efficiency of the sand filters located upstream the UV disinfection unit. Moreover, the *E. coli* influent concentration model (Foschi et al., 2021) estimated an increase in bacteria influent concentrations. Consequently, the developed UV disinfection model correctly estimates a probable

increase of effluent *E. coli* concentration and a positive PONC, even if the disinfection line worked at maximum number of active UV lamp banks. Moreover, it can be noted that the UV disinfection model estimates an increase in effluent *E. coli* concentration around August 26th, when another rain event occurred, but unfortunately no samples were collected. This is an example of how the model can support the management of UV disinfection when no data are available, providing in real-time continuous estimates of *E. coli* residuals in critical conditions, avoiding the time lag that is intrinsic of laboratory microbiological analyses.

It is important to stress one assumption under this use of the model: the regression models built in this work predict the disinfection efficiency statically, by linking current flow rate, UV dose and quality to the *E. coli* output and neglecting the transient dynamics deriving from the reactor hydrodynamics. However, this assumption is considered here as reasonable, since the hydraulic retention time of the disinfection contact channels is low, as typical of UV disinfection, ranging between 7 and 46 seconds.

4.3.3 Simulation of UV disinfection control for agricultural reuse

The WWTP UV disinfection facility under study was originally designed in 2004 to meet Italian requirements for agricultural reuse even in worst case scenarios. However, wastewater flow rate to be treated for reuse was not stationary over the years and even within the single irrigation seasons, as a consequence of a changing water demand by agriculture. Wastewater quality is not constant as well, as a result of inevitable fluctuations in raw sewage quality and upstream treatment efficiency. Then, the number of operating UV lamps could be adjusted according to these boundary conditions, in order to save energy, while maintaining adequate disinfection efficiency without overly cautious and blind safety factors. In order to demonstrate this approach, the TPANN UV disinfection model was used to simulate the control of the disinfection line in order to meet Italian requirements for agricultural reuse. Italian regulation states that the 10 CFU/100 mL limit referred to *E. coli* has to be met in 80% of samples collected in a year and *E. coli* residual concentration must never exceed 100 CFU/100 mL. Assuming that the plant utility must collect one control sample per day, it is possible to model each control event as a Bernoulli random variable, whose possible outcomes can be “compliance” or “non-compliance” with the discharge limit. Thus, the probability of being compliant k times on n control events can be described by a binomial probability function. Starting from these assumptions and since compliance is verified daily (n=356), in order to be compliant at least in 80% of the cases (k=292), the maximum probability of non-compliance for each control sample has to be 18%. Thanks to the bootstrap averaging approach, the TPANN model can estimate the effluent *E. coli* concentration as a distribution, incorporating model parameters uncertainty, and

can relate to regulation limits expressed as maximum PONC. Thus, a robust control strategy was implemented, which set at each time step (10 minutes) the optimal number of active UV lamp banks as the lowest guaranteeing the PONC setpoint (18%), with the constraint of non-exceeding the 100 CFU/100 mL limit. The estimated *E. coli* residual concentration in the effluent with optimal management of UV lamp banks resulting from the TPANN model-based control is reported in Figure 5.c. As shown, PONC is always lower than the desired limit of 18%, since the studied disinfection unit works only at three discrete level of UV dose, linked to the number of active UV lamp banks. Two active UV lamp banks per channel resulted as sufficient most of the time, while the use of all the UV lamp banks is necessary during some critical events, which in fact were denoted by an increase in wastewater turbidity and occurrence of rain events. The application of the control algorithm during the agricultural season would have implied a 63% saving of electric energy with respect to the business-as-usual policy of always using the maximum number of UV lamp banks, which maximizes system efficiency, but completely neglects the energy consumption minimization goal. Considering lamp nominal power consumption (about 45 kW per lamp bank), energy saving was estimated in about 580 MWh. During the 2018 irrigation season (June-August) energy consumption of UV disinfection accounted for about 12% of the WWTP total energy consumption, being the second most important item after biological oxidation: the control of the process could reduce at 5% the energy consumption of UV disinfection during the irrigation season, without any infrastructural intervention. Considering the average emission intensity of fossil CO_2 in Italy in 2018 (open-source data by European Environment Agency), this implies an avoided emission of roughly 145 tons of fossil CO_2. If the WWTP manager wants to apply a safety factor in designing the control algorithm, lower probability of non-compliance can be considered, at the price of a higher energy consumption. The trade-off between system reliability and energy consumption goals is reported in Figure 6, where all the equally optimal control alternatives are mapped in terms of the two conflicting objectives. It is evident from this relationship how it would be convenient to control UV lamp banks with a PONC setpoint lower than the minimum 18%. For example, working at 5% PONC, the energy saving would only slightly decrease, being about 58%. These results highlighted how exploring all equally optimal alternatives allowed to understand that, in the present case study, system reliability can be significantly increased at a small price in terms of energy saving.

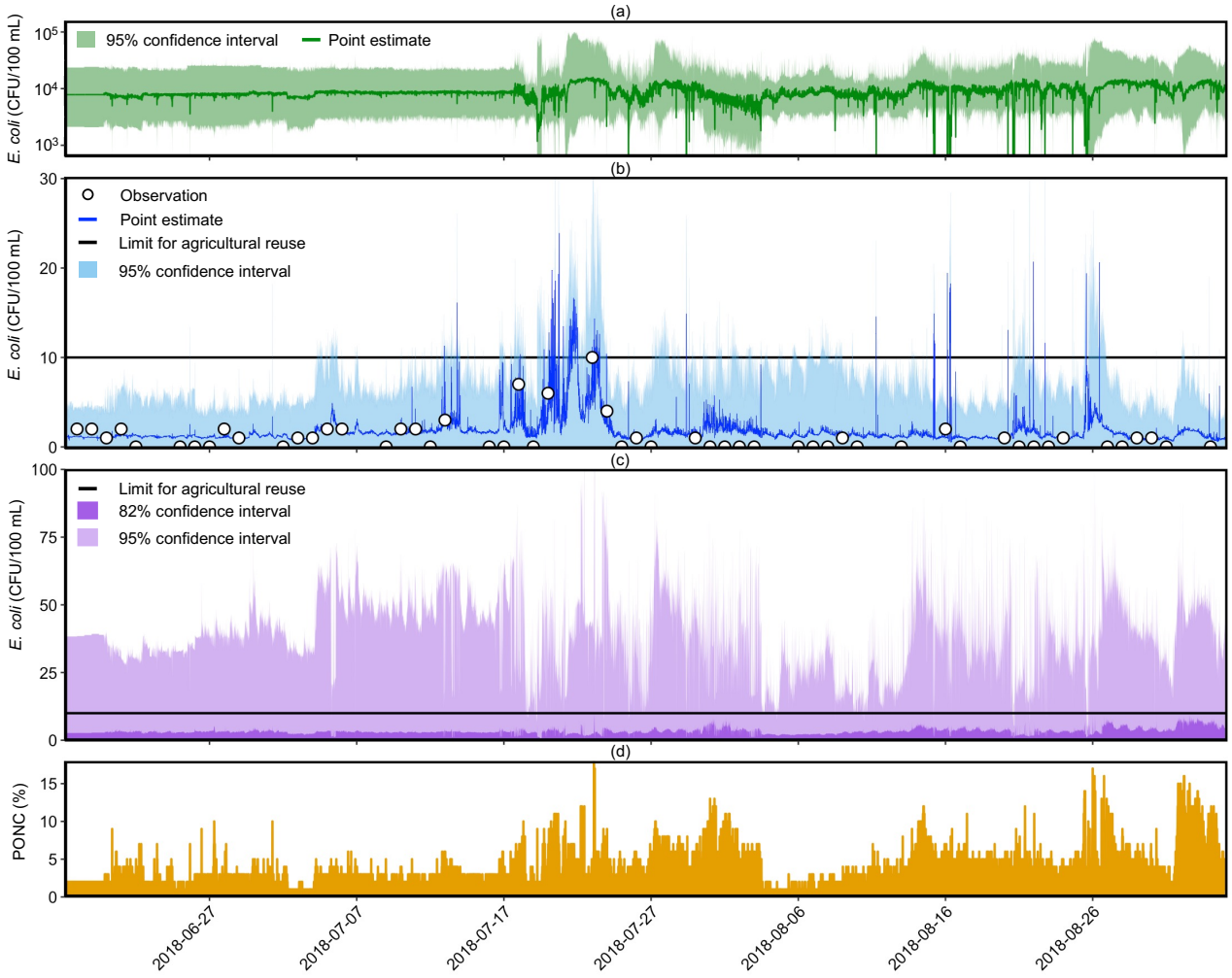

Figure 5 – Use of the two-part artificial neural network (TPANN) model during the agricultural season (June-August 2018): (a) prediction of influent *E. coli* concentration; (b) test of the TPANN model in predicting effluent *E. coli* concentration (disinfection operating at maximum number of UV banks; simulation of effluent *E. coli* concentration (c) and probability of non-compliance (d) resulting from control of UV lamp.

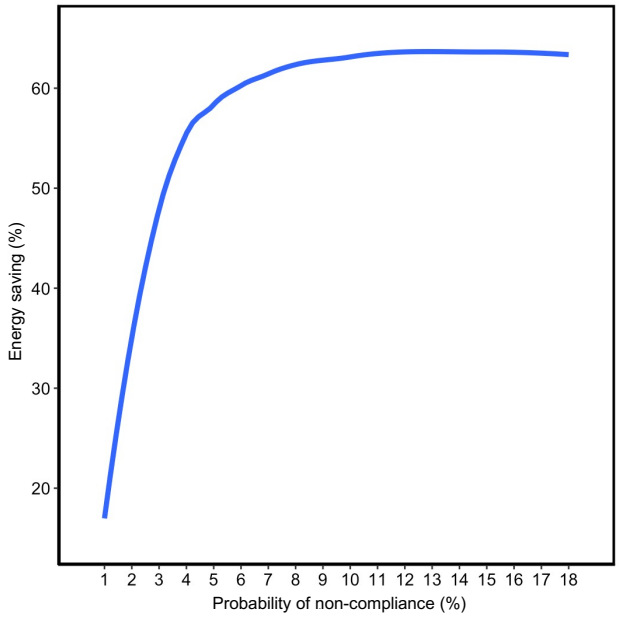

Figure 6 – Pareto front of energy saving and PONC from optimal control of UV lamp banks.

4.4 Conclusions

An ANN based "black-box" model of a full-scale wastewater UV disinfection facility was calibrated and tested on "unseen" data over a three-month time window. The model, named here as TPANN, proved good performances and was able to correctly predict even slight increases of a few CFU/100 mL in *E. coli* concentration, which is fundamental when dealing with high efficiency UV disinfection unit which has to meet the strict regulatory limits for wastewater reuse. The model was built as a "two-part" structure made of a sequence of an ANN classifier and an ANN regressor, which allowed both to handle the high proportion of zeros in the dataset and to catch nonlinear relationships in data. Such model structure could then be helpful in modeling any high efficiency disinfection unit, being able to use as calibration data all zero count events, without losing useful information to understand the behavior of the disinfection process at high disinfectant doses. Comparison of the TPANN model with several linear regression techniques supported the conclusion that a nonlinear family of models is needed to catch the relationship between number of UV lamps, flow rate, wastewater quality and disinfection efficiency and thus improve predictive performance. The TPANN model was here proposed as an alternative to traditional CFD-based modeling of UV disinfection, avoiding long computational time and thus unlocking real-time control applications. Real-time control of the operating banks of lamps of the UV disinfection facility under study was then simulated during the irrigation season of 2018 (from June to August), when the WWTP was delivering treated wastewater to an agricultural irrigation channel. Results highlighted that a potential saving of 63% of energy, corresponding to 7% of the total consumption of the WWTP, could be saved with respect to the conservative business-as-usual lamps management, since the model-based control allows to adjust the number of operating lamps according to the current flow rate and wastewater quality. Future research is needed to experiment and validate the model-based control of UV disinfection. Importantly, a dataset of adequate size should be collected to assess model accuracy in estimating the PONC. Moreover, important additional aspects of the process such as lamps fouling and the impact of lamp control on their durability should be addressed.

Acknowledgements

We would like to thank MM S.p.A., which manages the integrated water service in the city of Milan (Italy), and specifically Marco Blazina and Andrea D'Anna for having made this study possible. The authors would like to thank also all the plant engineering and technical team who supported the laboratory and field work: Chiara Pagano, Pasquale Cassatella, Katia Agnese Piva.

Last but not least, authors are grateful to Stefania Masyutina for their valuable support in field data collection.

References

Abu-ghararah, Z.H., 1994. Effect of Temperature on the Kinetics of Wastewater Disinfection Using Ultraviolet Radiation. J. Environ. Sci. Heal. . Part A Environ. Sci. Eng. Toxicol. 29, 585–603. https://doi.org/10.1080/10934529409376056

Ahmed, Y.M., Ortiz, A.P., Blatchley, E.R., 2019. Stochastic Evaluation of Disinfection Performance in Large-Scale Open-Channel UV Photoreactors. J. Environ. Eng. (United States) 145. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001562

Bolton, J.R., Linden, K.G., 2003. Standardization of Methods for Fluence (UV Dose) Determination in Bench-Scale UV Experiments. J. Environ. Eng. 129, 209–215. https://doi.org/10.1061/(asce)0733-9372(2003)129:3(209)

Bolyard, S.C., Motlagh, A.M., Lozinski, D., Reinhart, D.R., 2019. Impact of organic matter from leachate discharged to wastewater treatment plants on effluent quality and UV disinfection. Waste Manag. 88, 257–267. https://doi.org/10.1016/j.wasman.2019.03.036

Breiman, L., 1996. Bagging Predictors, URL: https://link.springer.com/article/10.1007%2FBF00058655. Mach. Learn. 24, 123–140. https://doi.org/10.1007/BF00058655

Carré, E., Pérot, J., Jauzein, V., Lopez-Ferber, M., 2018. Impact of suspended particles on UV disinfection of activated-sludge effluent with the aim of reclamation. J. Water Process Eng. 22, 87–93. https://doi.org/10.1016/j.jwpe.2018.01.016

Chahal, C., van den Akker, B., Young, F., Franco, C., Blackbeard, J., Monis, P., 2016. Pathogen and Particle Associations in Wastewater: Significance and Implications for Treatment and Disinfection Processes. Adv. Appl. Microbiol. 97. https://doi.org/http://dx.doi.org/10.1016/bs.aambs.2016.08.001

Chen, J., Deng, B., Kim, C.N., 2011. Computational fluid dynamics (CFD) modeling of UV disinfection in a closed-conduit reactor. Chem. Eng. Sci. 66, 4983–4990. https://doi.org/10.1016/j.ces.2011.06.043

Colin, C.A., Pravin, T., 2013. Regression analysis of count data, Second edition, Regression Analysis of Count Data, Second Edition. https://doi.org/10.1017/CBO9781139013567

Cragg, J.G., 1971. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. Econometrica 39, 829. https://doi.org/10.2307/1909582

Efron, B., Hastie, T., 2016. Computer age statistical inference: Algorithms, evidence, and data science, Computer Age Statistical Inference: Algorithms, Evidence, and Data Science. https://doi.org/10.1017/CBO9781316576533

Fenner, R.A., Komvuschara, K., 2005. A New Kinetic Model for Ultraviolet Disinfection of Greywater. J. Environ. Eng. 131, 850–864. https://doi.org/10.1061/(asce)0733-9372(2005)131:6(850)

Hastie, T. et. all., 2009. Springer Series in Statistics The Elements of Statistical Learning. Math. Intell. 27, 83–85. https://doi.org/10.1007/b94608

Hornik, K., Stinchcombe, M., White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Networks 3, 551–560. https://doi.org/10.1016/0893-6080(90)90005-6

Jolis, D., Lam, C., Pitt, P., 2001. Particle Effects on Ultraviolet Disinfection of Coliform Bacteria in Recycled Water. Water Environ. Res. 73, 233–236. https://doi.org/10.2175/106143001x139218

Lambert, D., 1992. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics 34, 1–14. https://doi.org/10.1080/00401706.1992.10485228

Lazarova, V., Savoye, P., Janex, M.L., Blatchley III, E.R., Pommepuy, M. 1999. Advanced wastewater disinfection technologies: State of the art and perspectives. Water Sci. and Techn. 40, 203-213. https://doi.org/10.1016/S0273-1223(99)00502-8

Lin, C.H., Yu, R.F., Cheng, W.P., Liu, C.R., 2012. Monitoring and control of UV and UV-TiO 2 disinfections for municipal wastewater reclamation using artificial neural networks. J. Hazard. Mater. 209–210, 348–354. https://doi.org/10.1016/j.jhazmat.2012.01.029

Madge, B.A., Jensen, J.N., 2006. Ultraviolet Disinfection of Fecal Coliform in Municipal Wastewater: Effects of Particle Size. Water Environ. Res. 78, 294–304. https://doi.org/10.2175/106143005x94385

Powell, C., Lawryshyn, Y., 2015. A method for determining the optimal discretization of UV lamps for emission-based fluence rate models. Water Sci. Technol. 71, 1768–1774. https://doi.org/10.2166/wst.2015.101

Severin, B.F., Suldan, M.T., Engelbrecht, R.S., 1983. Effects of Temperature on Ultraviolet Light Disinfection. Environ. Sci. Technol. 17, 717–721. https://doi.org/10.1021/es00118a006

Shu, C., Burn, D.H., 2004. Artificial neural network ensembles and their application in pooled flood frequency analysis. Water Resour. Res. 40, 1–10. https://doi.org/10.1029/2003WR002816

Talebizadeh, M., Belia, E., Vanrolleghem, P.A., 2014. Probability-based design of wastewater treatment plants. Proc. - 7th Int. Congr. Environ. Model. Softw. Bold Visions Environ. Model. iEMSs 2014 4, 2317–2324.

Wols, B.A., Hofman-Caris, C.H.M., Harmsen, D.J.H., Beerendonk, E.F., van Dijk, J.C., Chan, P.S., Blatchley, E.R., 2012. Comparison of CFD, Biodosimetry and Lagrangian Actinometry to Assess UV Reactor Performance. Ozone Sci. Eng. 34, 81–91. https://doi.org/10.1080/01919512.2012.651398

Xu, C., Rangaiah, G.P., Zhao, X.S., 2015. Application of Artificial Neural Network and Genetic Programming in Modeling and Optimization of Ultraviolet Water Disinfection Reactors. Chem. Eng. Commun. 202, 1415–1424. https://doi.org/10.1080/00986445.2014.952813

Environment Agency (2018). https://www.eea.europa.eu/data-and-maps/daviz/co2-emission-intensity-6#tab-googlechartid_googlechartid_googlechartid_googlechartid_chart_11111 (accessed 13th February 2021)

Chapter 5: Disinfection efficiency prediction under dynamic conditions: application to peracetic acid disinfection of wastewater

Abstract

In this work, a mechanistic dynamic model of continuous flow peracetic acid (PAA) disinfection was developed, calibrated and validated, assuming *E. coli* as indicator microorganism. The model was conceived as a 1-dimensional dispersion model integrating PAA first order decay and *E. coli* inactivation rate. Lab-scale batch experiments of PAA decay and *E. coli* inactivation experiments were performed to calibrate corresponding kinetic models. In each sample, conventional wastewater quality parameters were monitored. A PAA pilot reactor was set up to perform both tracer studies, for dispersion model calibration, and continuous flow disinfection experiments, to validate the integration of hydraulics and kinetics models, under both stationary and dynamic conditions. Linear regression models were calibrated to predict hydrodynamic dispersion, given the flow rate, and PAA decay parameters, given effluent quality and PAA dosage.

Successful validation of the PAA disinfection model proved the importance of (i) considering the disinfection process as a dynamic system and (ii) integrating real-time estimation of process disturbances, being the initial *E. coli* concentration and the impact of effluent quality and PAA dosage on PAA decay kinetics. Importantly, novel inactivation models were proposed, as two different modifications of a literature model for thermal inactivation. These models are suitable for dynamic simulation of Eulerian models and can describe the typical triphasic behavior of inactivation kinetics.

This chapter was submitted for publication.

5.1 Introduction

Chemical disinfection is one of the most widespread solutions for pathogen removal from wastewater, guaranteeing high inactivation efficiencies for bacteria, viruses and protozoa, while requiring simple and cheap facilities and equipment which are easy to be managed (White, 2010). Main drawbacks of most of chemical disinfectants are the formation of toxic disinfection by-products (DBPs) and a relevant ecotoxicological impact on natural waters due to both residual disinfectant and DBPs (*inter alia*: Yang et al., 2013, Du et al., 2017, How et al., 2017). Therefore, disinfectant dosage has to be optimized in order to guarantee the required pathogen reduction while minimizing residual concentration of disinfectant and process cost. In wastewater treatment, this goal is challenged by fluctuations in quality and flow rate of wastewater streams to be disinfected, which affect process efficiency. This variability is mainly due to the daily variability of raw wastewater quality and flow rate, and of the efficiency of upstream treatments. The change of wastewater quality impacts on decay kinetics of disinfectants (Wang et al., 2019, Domínguez Henao et al., 2018, Domínguez Henao et al., 2018), while flow rate variations cause different hydrodynamics in the disinfection contact tank. These dynamic conditions impact the two fundamental operating parameters in determining process efficiency: the concentration of disinfectant and contact time. Design, optimization and control of chemical disinfection require then a model of the process, which account for the impact of changing wastewater quality and flow rate when estimating pathogen inactivation.

One of the first systematic approaches to disinfection modeling was the Surface Water Treatment Rule (SWTR) and it was conceived for a highly conservative design of disinfectant dosage (USEPA, 2006). The SWTR assumes that the "dose" (CT), defined as the product of disinfectant concentration and residence time in the contact tank and expressed as mg L^{-1} min, is the primary factor determining disinfection efficiency, and reported CT tables for a variety of disinfectants, linking CT values to inactivation levels. The assumption on the importance of the dose is the basis of most of the following modeling approaches to disinfection, but still the SWTR is too conservative in the way the dose was estimated. The SWTR assumes that the retention time at which 10% of the water passes through the contact tank (T_{10}) and the residual concentration of disinfectant after the average hydraulic retention time have to be considered to compute the CT value, without accounting for specific characteristics of hydraulics and kinetics of the case study. This is the consequence of a single-objective approach, which look for a high safety factor when considering pathogen inactivation, while neglecting the environmental and economic impact of unnecessary overdosage. The modeling approach introduced by the Integrated Disinfection Design

Framework (IDDF) (Ducoste et al., 2001) represented a fundamental shift towards a site-specific modeling of continuous flow disinfection. IDDF integrates sub-models of contact tank hydraulics, disinfectant decay, DBPs formation and microbial inactivation kinetics, which are all calibrated on the case study of interest, allowing to optimize dosage while considering the specificity of disinfectant, wastewater and reactor characteristics. However, IDDF was primarily developed for drinking water and considers disinfection as a stationary process, with constant water quality and flow rate, which makes it very approximate for realistic wastewater applications.

More recently, variability of disturbances and model uncertainty were included in wastewater disinfection modeling, leading to more accurate and robust predictions of process efficiency. A first group of studies proposes stochastic models, which characterize the variability of process inputs and parameters as probability distributions (*inter alia*: Neumann et al., 2007; Santoro et al., 2015). This kind of approaches provide useful tools to deal with disturbances variability and model uncertainty, but they do not explicitly explain how time variability of wastewater quality and flow rate impact on the dynamic behavior of the disinfection process. In fact, variability and uncertainty are described as static probability distributions and propagated onto disinfection efficiency predictions by Monte Carlo simulations of the disinfection reactor under stationary conditions. Manoli et al. (2019) made a fundamental step forward in predicting the temporal dynamics of chemical disinfection, by modeling the process over time as succession of stationary conditions. The contact tank was modeled as a Tank-in-Series (TIS) model (Levenspiel, 1999), a series of ideal continuously stirred tank reactors. At each time step, residual peracetic acid (PAA) and *E. coli* concentrations were estimated with the Segregated Flow Model (Crittenden et al., 2017), using the current value of flow rate and PAA concentration at the reactor inlet. Elhalwagy et al. (2021) developed a disinfection model based on computational fluid dynamics to describe contact tank hydraulics and adding the description of the impact of solid settling on disinfectant decay and microbial inactivation. Despite these recent advances, wastewater disinfection was rarely modeled as a dynamic process and still a dynamic model of disinfection needs to be validated.

Statistical regression models represent an effective alternative to previously described mechanistic approaches whenever lack of knowledge on specific phenomena affecting disinfection can be compensated by directly capturing input-output relationships in data. For example, regularized linear models (*inter alia*: Kadoya et al., 2020, Kadoya et al., 2021) and neural networks (Newhart et al., 2021) were used to predict process efficiency and/or residual concentration of disinfectant. Moreover, data driven models, like neural networks, were used to build surrogate models of their mechanistic counterparts (Wei et al., 2020), with much lower computational time, unlocking applications in optimization and control.

This work aims at conceiving, calibrating and validating a mechanistic dynamic model of PAA disinfection. The process was described by a 1-dimensional advection-diffusion-reaction model, integrating PAA decay and inactivation kinetics. Impact of varying wastewater quality and flow rate on kinetic and hydraulic parameters was estimated by means of regression models. PAA decay and inactivation kinetic models were calibrated in lab-scale batch experiments. *E. coli* was used as indicator microorganism. A statistical regression model was calibrated to predict the impact of changing wastewater quality on PAA decay kinetics. The 1-dimensional advection-diffusion transport model was calibrated over tracer studies carried out on a pilot reactor. Finally, the overall 1-dimensional advection-diffusion-reaction model, integrating kinetic and hydraulic models, was validated over pilot-scale continuous flow disinfection experiments, both in stationary and dynamic conditions.

5.2 Materials and Methods

5.2.1 Model development

A mechanistic model of PAA disinfection was developed, integrating the description of contact tank hydraulics (paragraph 2.1.1), kinetics of PAA decay (paragraph 2.1.2) and kinetics of *E. coli* inactivation by PAA (paragraph 2.1.3). The PAA disinfection model was then calibrated and validated at pilot-scale under both stationary flow rate conditions (paragraph 2.1.4), and dynamic conditions (paragraph 2.1.5).

5.2.1.1 Contact tank hydraulics

A 1-dimensional advection-diffusion-reaction model, also known as *dispersion model* (Levenspiel, 1999), was used to describe fate and transport of reactive substances through the contact tank:

$$\frac{dS}{dt} = -v\frac{dS}{dx} + D\frac{d^2S}{dx^2} + \left(\frac{dS}{dt}\right)_r \qquad \text{(Eq. 2.1)}$$

where S (mg L^{-1}) is the concentration of a generic reactive substance, x (m) is the spatial coordinate indicating the distance from the inlet of the contact tank, v (m min^{-1}) is the flow velocity along x and D (m^2 min^{-1}) is the dispersion coefficient. $\left(\frac{dS}{dt}\right)_r$ (mg L^{-1} min^{-1}) is the rate of the reaction involving S. The model assumes that S is constant over the cross-sectional area of the reactor. The model does not explicitly describe the flow inversions due to the *chicane* of the contact tank and considers it as a linear channel. The additional mixing and dispersion due to flow inversions is then incorporated in D.

Two models were compared for contact tank hydraulics, being: (i) a single dispersion model (SDM) (Figure 1.a), describing the contact tank as a single linear channel, and (ii) the parallel combination of two dispersion models (PDM), describing the tank as two parallel linear channels with a final blend of flow rates (Figure 1.b). In this second case, total flow rate and cross-sectional area are split between the two dispersion models according to two parameters α and β, respectively, with $\alpha, \beta \in [0,1]$.

Equations describing the response of the dispersion model to an impulse of a non-reactive substance under open-channel boundary conditions were used to calibrate D on tracer studies data (Crittenden et al., 2017):

$$E(\vartheta) = \frac{V}{M} C_{pulse} = \begin{cases} \frac{1}{\sqrt{4\pi\vartheta d}} e^{-\frac{(1-\theta)^2}{4\vartheta d}}, & d > 0.01 \\ \frac{1}{\sqrt{4\pi d}} e^{-\frac{(1-\theta)^2}{4d}}, & d \leq 0.01 \end{cases} \quad \text{(Eq. 2.2)}$$

$$d = \frac{D}{vL} \quad \text{(Eq. 2.3)}$$

$$\theta = \frac{t}{HRT} \quad \text{(Eq. 2.4)}$$

where M is the mass of the injected tracer, V (m^3) is the contact tank volume, C_{pulse} (mg L^{-1}) is the concentration of tracer leaving the contact tank at time t (min), d is the *dispersion number* (-), which measure the axial dispersion of the contact tank, HRT (min) is the average hydraulic retention time in the contact tank, L (m) is the length of the path along x from the inlet to the outlet of the contact tank. $E(\vartheta)$ is a dimensionless function known as residence time distribution (Levenspiel, 1999). The SDM impulse response is directly described by Eq. 2.2, the PDM impulse response is the result of the combination of the two channels:

$$E(\vartheta) = \alpha E_1(\vartheta) + (1-\alpha) E_2(\vartheta) \quad \text{(Eq. 2.5)}$$

$E(\vartheta)$ function was calibrated on tracer experiments (see paragraph 2.2) at different flow rates, considering one at a time datasets from each tracer experiment and getting a different parametrization for each flow rate value. Empirical models of D, α and β as linear, power and exponential functions of flow rate were then calibrated.

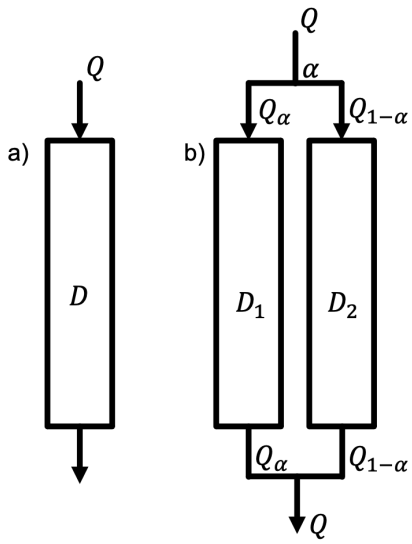

Figure 1 – Schemes of the two options for the hydraulic model of the pilot contact tank.

5.2.1.2 PAA decay

PAA decay was modeled with the Haas and Finch model (Haas and Finch, 2001), which implies first order decay rate (k) and instantaneous oxidative demand (OD) after dosage:

$$\frac{dPAA}{dt} = -kPAA \quad \text{(Eq. 2.6)}$$

$$PAA(t) = (PAA - OD)e^{-kt} \quad \text{(Eq. 2.7)}$$

with PAA (mg L^{-1}) as PAA concentration, PAA_0 (mg L^{-1}) as PAA dosage, and k (min^{-1}) and OD (mg L^{-1}) as model parameters to be calibrated.

Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996) was used to identify and calibrate linear models to predict k and OD in the effluent, given monitored conventional wastewater quality parameters as potential predictors (see paragraph 2.2). In LASSO regression, optimal coefficients of a linear model are found by minimizing a penalized residual sum of squares plus a penalty proportional to the sum of the absolute values of coefficients:

$$SSE^{LASSO} = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p}|\beta_i| \quad \text{(Eq. 2.8)}$$

where y_i are the observations, $\hat{y}_i$ are the predicted values, N is the size of the sample and β_j are model parameters. LASSO has the advantage of shrinking continuously towards zero coefficients of irrelevant predictors during SSE^{LASSO} minimization, thus leading to a sparse model with a reduced number of predictors. The value of λ was tuned by minimizing Mean Squared Error (MSE) within a Leave-one-out Cross-validation (LOOCV) procedure (Hastie, 2009). Then, the final model was calibrated on the whole dataset by LASSO using the optimal value of λ.

5.2.1.3 E. coli inactivation

Four models were compared for *E. coli* inactivation: (i) Dose Model (DM) (Domínguez Henao et al., 2018b); (ii) Double Exponential Model (DEM) (Santoro et al., 2015); (iii) and (iv) two modified versions of the Geeraerd inactivation model (MGM1 and MGM2) (Geeraerd et al., 2000).

The DM describes inactivation by PAA in batch conditions as:

$$log\frac{N}{N_0} = -\frac{k'D^n}{1+e^{h-D}} \quad \text{(Eq. 2.9)}$$

$$d(t) = \int_0^t C dt = \left(\frac{C_0 - OD}{k}\right)(1 - e^{-kt}) \quad \text{(Eq. 2.10)}$$

where N_0 and N (CFU/100 mL) are the initial concentration of bacteria and the concentration at time t, and k', n and h are model parameters. $D(t)$ (mg L^{-1} min) is the disinfectant dose at time t, i.e. the integral of PAA concentration up to time t. This definition of the dose synthesizes the combined variation of PAA concentration and contact time, while accounting for the progressive loss of PAA concentration due to decay. DM can effectively describe with only 3 parameters the triphasic behavior of PAA inactivation kinetics, made of a first *lag,* followed by a sequence of a log-linear phase with high inactivation rate and, finally, by a sharp decrease of inactivation rate phase (*tailing*).

The DEM used in this work is inspired to previously developed literature model (Santoro et al., 2015), postulating that the high-rate log-linear phase of inactivation is mainly due to the rapid inactivation of dispersed *free-swimming* bacteria, while the final tailing with slower rate is caused by the shielding effect that protects bacteria which aggregate on suspended solid particles. The total population of bacteria in the undisinfected effluent is then assumed to be divided in *free-swimming* and *particle-associated* bacteria. In this work, to define DEM both in a derivative and in an integral form, *free-swimming* and *particle-associated* bacteria are assumed to follow a Chick-Watson inactivation kinetics, as described by:

$$\frac{dN_{fs}}{dt} = -aN_{fs}PAA \quad \text{(Eq. 2.11)}$$

$$\frac{dN_{pa}}{dt} = -bN_{pa}PAA \quad \text{(Eq. 2.12)}$$

where N_{fs} (CFU/100mL) is the concentration of *free-swimming* bacteria and N_{pa} (CFU/100mL) is the concentration of *particle-associated* bacteria

Relying on the assumption that the total bacteria population can be calculated as the sum of the two above-mentioned subpopulations, the overall DEM kinetics in its integral form can be derived as:

$$N = N_{fs} + N_{pa} = N_{0,fs}e^{-ad} + N_{0,pa}e^{-bd} = \delta N_0 e^{-ad} + (1-\delta)N_0 e^{-bd} \quad \text{(Eq. 2.13)}$$

where δ is the percentage of N_{fs} in the total concentration of initial bacteria (N_0). The model parameters are a, b and δ.

As for MGM1 and MGM2, the first one is based on the same assumption as DEM on the presence of *free-swimming* and *particle-associated* bacteria, but it assumes a different inactivation kinetics for the first group, that can describe the initial lag, under the hypothesis that the lag is due to the presence of components (C) protecting bacteria from inactivation (Geeraerd et al., 2000). Here it is then assumed that C undergoes a second order inactivation dependent on PAA concentration. MGM1 is then formulated as:

$$\frac{dN_{fs}}{dt} = -\eta N_{fs} PAA \quad \text{(Eq. 2.14)}$$

$$\eta = \eta_{max}\left(1 - \frac{C}{C-K_c}\right) = \eta_{max}\frac{1}{1+X} \quad \text{(Eq. 2.15)}$$

$$X = \frac{C}{K_c} \quad \text{(Eq. 2.16)}$$

$$\frac{dX}{dt} = -\eta_{max} PAA X \quad \text{(Eq. 2.17)}$$

where the Michaelis-Menten based kinetics η is introduced in the inactivation rate of bacteria. K_c (units cell^{-1}) stands for the value of C (units cell^{-1}) where η equals half of its final value, which is 1. X can be interpreted as a measure of the physiological state of the population (Geeraerd et al., 2000). The integral form of Eq. 2.14 and of the MGM1 kinetics are given respectively by:

$$\frac{N_{fs}}{N_{fs,0}} = \frac{X_0+1}{X_0+e^{\eta max d}} \quad \text{(Eq. 2.18)}$$

$$N = \frac{\delta N_0 (X_0+1)}{X_0+e^{\eta max d}} + (1-\delta) N_0 e^{-bd} \quad \text{(Eq. 2.19)}$$

MGM2 was defined as MGM1, except for the decay of X (and, consequently, C), which was described as a first order decay, not depending on PAA concentration:

$$\frac{dX}{dt} = -\eta_{max} X \quad \text{(Eq. 2.20)}$$

MGM2 cannot be integrated analytically and was then solved numerically.

5.2.1.4 Prediction of residual E. coli and PAA concentrations under stationary flow rate conditions

Residual *E. coli* and PAA concentration at the outlet of the pilot contact tank during continuous flow disinfection at constant flow rate were predicted with the SFM (Levenspiel, 1999). The SFM works under the assumption that all the fluid elements are segregated, namely they do not mix or interact with each other. Therefore, the amount of any reaction that takes place in each fluid element is estimated and then the elements are mixed at the end of the reactor. Usually, every fraction of fluid is approximated by a batch reactor with the same contact time of that fluid fraction

retention time, identified by the $E(\vartheta)$ function. The prediction of residual bacteria and PAA concentrations by SFM can be then formalized as a convolution integral:

$$PAA_{out} = PAA_0 \int_0^\infty PAA_{batch}(t)E(t)dt \quad (2.21)$$

$$N_{out} = N_0 \int_0^\infty N_{batch}(t)E(t)dt \quad (2.22)$$

$$E(t) = \bar{t}E(\vartheta) \quad (2.23)$$

where $E(t)$ is the RTD function defined over time and $\bar{t}$ (min) is the average contact time. $PAA_{batch}(t)$ and $N_{batch}(t)$ are PAA and bacteria concentration trends, respectively, in batch conditions, defined by models described in paragraphs 2.1.2 and 2.1.3.

5.2.1.5 Prediction of residual E. coli and PAA concentrations under dynamic flow rate conditions

Residual *E. coli* and PAA concentration at the outlet of the pilot contact tank during continuous flow disinfection at variable effluent flow rate and water quality was simulated by numerical solution of the dispersion model in its derivative form. Fate and transport of bacteria and PAA through the contact tank were described as:

$$\frac{dN_{fs}}{dt} = -\frac{Q}{A}\frac{dN_{fs}}{dx} + D(Q)\frac{d^2N_{fs}}{dx^2} - \eta N_{fs}PAA \quad \text{(Eq. 2.24)}$$

$$\frac{dN_{pa}}{dt} = -\frac{Q}{A}\frac{dN_{pa}}{dx} + D(Q)\frac{d^2N_{pa}}{dx^2} - bN_{pa}PAA \quad \text{(Eq. 2.25)}$$

$$\frac{dPAA}{dt} = -\frac{Q}{A}\frac{dPAA}{dx} + D(Q)\frac{d^2PAA}{dx^2} - k(\bar{q})PAA \quad \text{(Eq. 2.26)}$$

Equations are reported assuming MGM1 as inactivation model, but can be straightforwardly derived for DEM or MGM2 by replacing the reaction term in equations 2.17 and 2.18. Dispersion parameter D is dependent on flow rate (Q), while PAA decay parameter k depends on wastewater quality parameters $(\bar{q})$ at time t. $\bar{q}$ corresponds to the group of water quality parameters which were selected as good predictors of k by LASSO regression (paragraph 2.1.2). In the case of PDM, Eq. 2.24-2.26 were solved for the two channels and concentrations flowing out the contact tank were computed as the final blend.

E. coli concentration at the inlet of the contact tank was predicted by a linear regression model, calibrated according to LASSO and LOOCV (paragraph 2.1.2), considering wastewater quality parameters as predictors (paragraph 2.2).

5.2.1.6 Experimental plan

The experimental plan was designed to calibrate kinetic models of PAA decay and inactivation of *E. coli* and the hydraulic model of the contact tank, and to validate the integration of such models

to predict PAA and *E. coli* residual concentration in a continuous flow disinfection process, both in stationary and dynamic flow rate conditions.

PAA decay and disinfection experiments in batch conditions were carried out in the laboratory on grab samples of the undisinfected effluent collected in Peschiera Borromeo municipal wastewater treatment plant (WWTP) in Milan urban area (Italy), between 02/09/2019 and 28/01/2020, during random moments of day and week. The WWTP (440.000 PE) includes a primary sedimentation stage, followed by a suspended biomass nitrification/oxidation and phosphorous removal by aluminium chloride. The stream quality is then refined by a double stage biofilter (BIOFOR®, Suez) performing nitrification, denitrification and suspended solids filtration.

Each collected sample was characterized by pH, conductivity (CND), temperature (T), total suspended solids (TSS), turbidity (TRB), chemical oxygen demand (COD), ultraviolet absorbance at 254 nm (UV254), total nitrogen (TN), ammonia (NH_4^+) and nitrate (NO_3^-).

Tracer experiments were carried out on the pilot contact tank at 5 flow rates (35, 60, 80, 100, 130 L min^{-1}). An impulse of 0.1 kg of NaCl was pumped at the inlet of the contact tank and conductivity was monitored over time at the outlet.

PAA disinfection experiments in continuous flow conditions were carried out at pilot-scale in between 14/11/2019 and 10/12/2021, under both stationary and dynamic flow rate conditions.

Experiments are summarized in Table 1, including details about data use in model calibration or validation.

Table 1 – Summary of the experimental plan. Sampling points refer to Figure 2.

ID	Type of experiment	Condition	Scale	Flow rate	Measured quantity	Number of experiments	Sampling points	Modeling phase
1	Tracer study	continuous	pilot	stationary	− conductivity	5	P1	calibration of hydraulic model (par. 2.1.1)
2	PAA decay	batch	lab	-	− PAA − wastewater quality	27	P1	calibration of PAA decay (par. 2.1.2)
3	*E. coli* inactivation	batch	lab	-	− *E. coli* − wastewater quality	7	P2	calibration of PAA inactivation (par. 2.1.3)
4	PAA disinfection	continuous	pilot	stationary	− PAA − *E. coli* − wastewater quality	9	P1; P2	validation of disinfection model (par. 2.1.5)
5	PAA disinfection	continuous	pilot	dynamic	− PAA − *E. coli* − turbidity − conductivity − pH − UV254	2	P1; P2	validation of disinfection model (par. 2.1.6)

5.2.1.7 Pilot plant

A pilot-scale disinfection plant was set up in the WWTP. It is a 2.2 m^3 open channel *chicane* type reactor, whose horizontal and vertical sections are reported and detailed in Figure 2. The plant inlet is equipped with an online propeller flow meter (Digiten FL-1608), a motor valve (Burkert 3285) and a peristaltic dosing pump (SEKO) for PAA dosage. All devices are connected to an Arduino UNO programmable board for flow rate readings, control of valve opening and dosing pump flow rate. The contact tank was fed on the undisinfected tertiary effluent of the WWTP, by a centrifugal pump (DAB) withdrawing from the inlet of the full-scale disinfection contact tank. The pilot plant enabled continuous flow disinfection experiments under constant flow rate or arbitrary flow rate patterns and real-time wastewater quality variations. Samples for batch (lab-scale) and continuous flow (pilot-scale) disinfection experiments and tracer studies were collected at inlet (P1) and outlet (P2) of the contact tank, as highlighted in Figure 2.

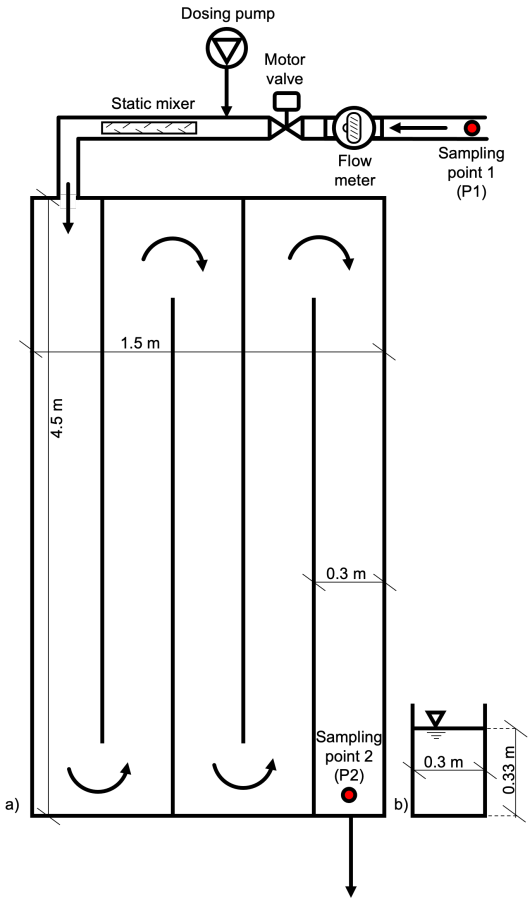

Figure 2 – Scheme of the disinfection pilot plant (not in scale).

5.2.1.8 Lab-scale decay and disinfection experiments

1-h PAA decay tests on effluent samples were performed in completely mixed batch reactors (CMBR) mixed by magnetic stirrer at room temperature (20 ± 1 °C). In each experiment, PAA

commercial solution (PROMOX P510, PAA content: 15% w/w, hydrogen peroxide (H_2O_2) content: 23-24% w/w) was added to the effluent to obtain the given PAA initial concentration. Sample aliquots were collected at five contact times (2, 5, 10, 30 and 60 min) to measure residual PAA concentration. Decay experiments were carried out on 9 samples collected at an initial PAA concentration of 1, 2 and 5 mg L^{-1}, for a total of 27 experiments, evaluating PAA residual concentration.

In disinfection experiments, carried out in the same setup of PAA decay tests, 7 undisinfected effluent samples were exposed to different combinations of PAA concentration (1 to 5 mg L^{-1}) and contact time (1 to 60 minutes), resulting in monotonically increasing disinfectant doses (1 to 270 mg L^{-1} min). *E. coli* concentration was measured in each sample before and after the exposure to each dose level.

5.2.1.9 Pilot-scale disinfection experiments

Continuous flow disinfection experiments were performed on the pilot contact tank, fed with the real WWTP undisinfected tertiary effluent and thus with variable effluent quality conditions. PAA was continuously dosed at constant concentration.

In stationary conditions experiments, flow rate was kept constant, while PAA and *E. coli* concentration was measured at the outlet after 4 times the average HRT. At the same time, a sample was collected at the inlet, where pH, CND, T, TSS, TRB, COD, UV254, TN, NH_4^+ and NO_3^- were measured. Stationary flow rate experiments were carried out in 9 combinations of PAA dosage (1.5, 3 and 4.5 mg L^{-1}) and flow rate (50, 70 and 100 L min^{-1}). PAA and flow rate values were chosen to be different from the calibration phase, but still avoiding extrapolation.

In dynamic conditions experiments, flow rate varied according to fluctuating patterns (40 to 140 L min^{-1}, with HRT ranging between 17 and 55 min), while PAA concentration was kept constant at 1 mg L^{-1} (1st experiment) and 3 mg L^{-1} (2nd experiment). PAA concentration at the outlet of the contact tank was measured every 10 minutes. *E. coli* were measured both at the inlet, every 60 minutes, and at the outlet of the contact tank, every 20 minutes. TRB, pH, CND and UV254 were measured each 10 minutes at the inlet. Experiments lasted 5 hours each.

5.2.2 Chemical and microbiological analyses

PAA residual concentration was measured by the DPD method according to Dominguez-Henao et al. (2018). Commercial test kits were used for measurement of COD (Hach LCI500, ISO 15705), ammonia (Hach LCK303, ISO 7150-1), nitrate (Hach LCK339, EN38405 D-2) and total nitrogen (Hach LCK238, EN ISO 11905-1). TSS, turbidity and UV254 were respectively measured by 0.45-

μm membrane filtration (Standard Methods, section 2540B, APHA/AWWA/WEF, 2012), portable turbidimeter (VELP Scientifica) and 1-cm optical path laboratory spectrophotometer (Hach DR6000). pH and conductivity were measured with a multiparameter probe (HACH HQ40d).

E. coli were enumerated by membrane filtration method according to Standard Methods (Section 9222, APHA/AWWA/WEF, 2012), using 0.45- μm pore size cellulose nitrate membranes (Whatman) and chromogenic agar (Microinstant® Chromogenic Coliforms Agar, Scharlau) as culture medium. Inoculated plates were incubated at 37 °C for 24 h. *E. coli* and TC were expressed as CFU in 100 mL volume.

5.2.3 Data processing

Models described in paragraphs from 2.1.1 to 2.1.4 were calibrated with the software R (version 4.0.1). Package *nls 0.2* was used for nonlinear regression and package *glmnet 4.0* was used for LASSO. Mathworks Matlab R2021b was used for simulations described in paragraphs 2.1.4 and 2.1.6. Matlab Partial Differential Equation Toolbox was used for numerical solution of ordinary differential equations for simulations of disinfection according to the dispersion model (Eq. 2.24 to 2.26).

5.3 Results and Discussion

Firstly, the three sub-models constituting the mechanistic model were calibrated (paragraphs 3.1 to 3.3) on data derived from tracer study and batch decay/inactivation tests. The calibrated models were then used to predict *E. coli* and PAA concentration at the outlet of the pilot reactor during continuous flow disinfection (paragraphs 3.4 and 3.6). For dynamic simulations, considering the time-varying flow rate and wastewater quality, the concentration of *E. coli* at the inlet of the pilot reactor was estimated in real-time.

5.3.1 Calibration of the hydraulic model

Model parameters of SDM and PDM (paragraph 2.1.1) were estimated by nonlinear regression using tracer tests data performed at different HRT, ranging between 17 and 56 minutes. Results of model calibrations are reported in Table S1. The PDM fitted experimental data much better (adjusted R^2: 0.861 to 0.961) than the SDM (adjusted R^2: 0.462 to 0.946), suggesting that the flow through the contact tank could be described as two parallel non-mixing fluxes with different longitudinal dispersion. PDM fit of experimental tracer data is reported in Figure 3a. This result is consistent with previous studies on the effectiveness of *compartment models* in describing hydrodynamics of chemical disinfection reactor, which highlighted that most of contact tanks are

best described by two TIS models (Gorzalski et al., 2018), which can be considered as equivalent of dispersion models (Levenspiel, 1999). PDM parameters suggest that the model is defined by a main "low-dispersion" flux (D_1=0.24-0.56) and a secondary "high-dispersion" flux (D_1=0.93-10.88). The lower the total flow rate, the higher the importance of the "low-dispersion" flux, as can be argued by values of parameter α (0.51-0.87) and β (0.40-0.89). Interestingly, dispersion in the "high dispersion" flux decreases with flow rate, while the opposite happens for the other flux. The variation of the estimated parameters with flow rate evidenced how velocity in the contact tank impacts on its overall hydrodynamic behavior, determining the degree of the shift from the ideal plug-flow model (Furman et al., 2005, Zheng & Mackley, 2008). Empirical linear models were calibrated to describe the relationships between flow rate and each PDM parameter. Model equations, coefficients of determination and statistical significances are reported in Table 2**,** while models fits are reported in Figures 3b-e The integration of these equations in the PDM makes it able to account in real-time for changes in hydrodynamic behavior of the contact tank due to flow rate fluctuations.

Table 2 – Equations, coefficients of determination and significance of the empirical models of the dispersion model parameters

Parameter	Equation	R^2	p-value
D_1	$D_1 = -0.003Q + 0.71$	0.965	0.003
D_2	$D_2 = 0.11Q - 5.49$	0.879	0.019
α	$\alpha = 0.004Q + 0.40$	0.748	0.058
β	$\beta = 0.006Q + 0.21$	0.888	0.016

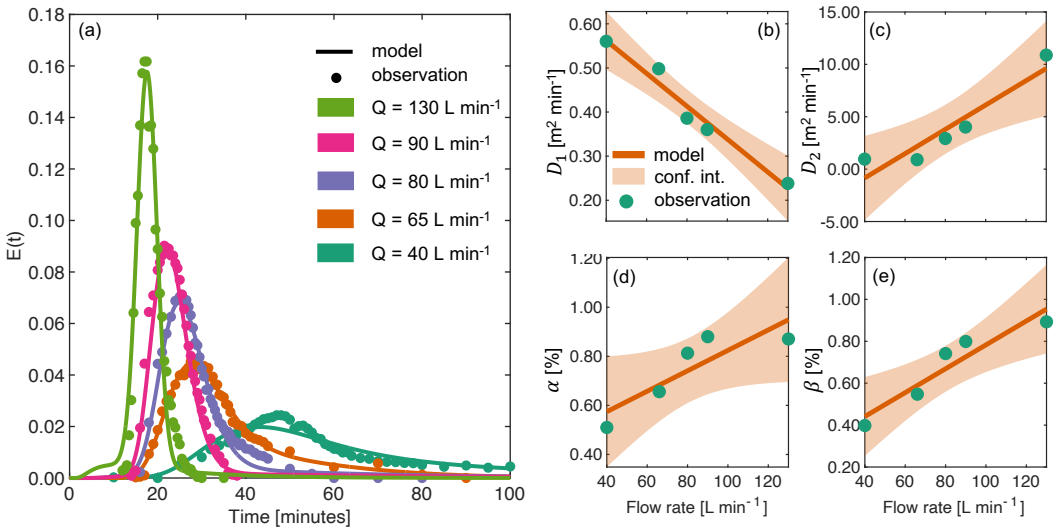

Figure 3 – Fitting of PDM on tracer experiments at different flow rates (a) and fitting of empirical models predicting PDM parameters at given flow rate (b).

5.3.2 Calibration of the PAA decay model

Data from PAA decay experiments evidenced high variability of both first order decay rate and oxidative demand. A single parametrization of the Haas and Finch model was not adequate for a sufficiently accurate prediction of residual PAA, as reported in Figure S1. Haas and Finch model was then calibrated on each out of 27 decay experiments. k and OD model fits are illustrated in Figure 4 and model coefficients and performances are reported in Table 3. LASSO regression revealed that part of this variability can be explained by two linear models using PAA dosage and conventional effluent quality parameters as predictors of k and OD. As a result of the LOOCV procedure, 4 variables (initial PAA dosage, TRB, UV254, TN and NO_3^-) were selected as predictors of k, as reported in Figure 4a, where MSE reaches its minimum when λ is such that only 4 predictor coefficients are non-zero. This result is consistent with recent findings on PAA. Higher PAA concentrations imply higher stability of the compound in water and lower decay, while concentration of suspended solids and soluble organics, which are well correlated to TRB and UV254, increase k (Santoro et al., 2015, Domínguez Henao al., 2018a, 2018b, Elhalwagy et al., 2021). A physical relationship between TN and NO_3^- and PAA decay has not been documented and these two parameters probably work here as proxies of the upstream activated sludge treatment operating conditions. As for OD, only PAA dosage was selected as predictor (Figure 4c). Despite some limitations, as the poor performance in OD prediction, both models were integrated in the disinfection model, in order to estimate in real-time the impact of changing wastewater quality and dosage on PAA decay.

Table 3 – Results of calibration and LOOCV for regression models predicting PAA first order decay rate (k), PAA oxidative demand (OD) and *E. coli* concentration at the inlet of the pilot reactor. "cal." (calibration): performances obtained by LASSO regression on the whole dataset at optimal λ. "CV" (cross-validation): performances by LASSO regression from LOOCV.

	$log_{10}k$		OD		$log_{10}E.coli$	
PAA_0	-0.041		0.073		-	
pH	-		-		0.049	
CND	-		-		-	
T	-		-		-	
TSS	-		-		-	
TRB	0.161		-		0.415	
COD	-		-		-	
UV254	0.029		-		-	
TN	-0.039		-		-	
NH^{4+}	-		-		-	
NO^{3-}	-0.034		-		-0.189	
Intercept	-2.159		0.164		4.699	
	cal.	**CV**	**cal.**	**CV**	**cal.**	**CV**
R2	0.67	0.45	0.29	0.10	0.67	0.58

MAE	0.003	0.004	0.135	0.149	0.385	0.502
MAPE	38%	49%	8.5%	9.3%	8.4%	11.2%

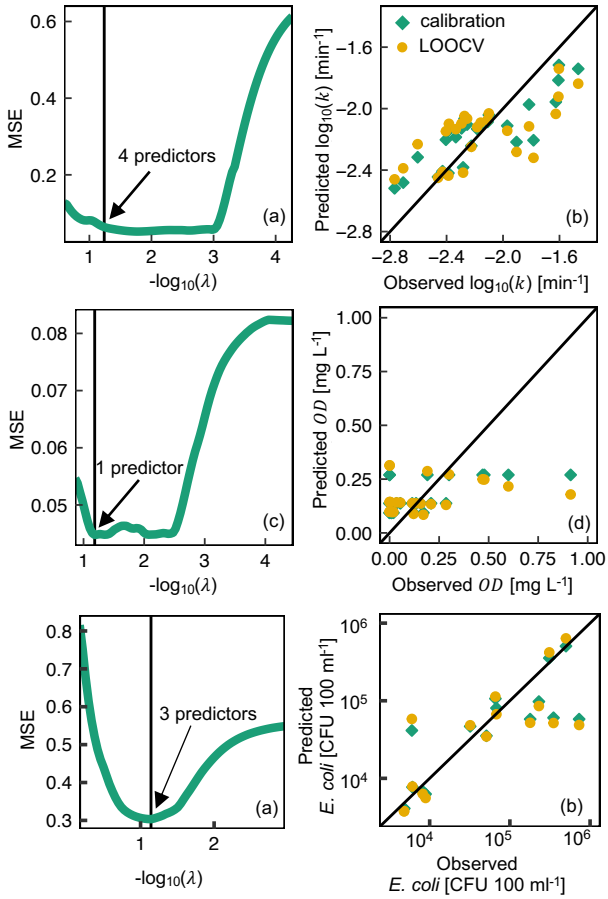

Figure 4 – Results of LASSO regression and LOOCV for models of PAA decay kinetic parameters (k and OD).

5.3.3 Calibration of the E. coli inactivation model

Data collected in lab-scale batch disinfection experiments, reported in Figure 5, revealed the typical triphasic trend of bacteria inactivation by PAA (*inter alia*: Rossi et al., 2007, Antonelli et al., 2013, Santoro et al., 2005, McFadden et al., 2017, Dominguez Henao et al., 2018). Experimental data suggest an initial lag phase with low inactivation rate for doses lower than approximately 10 mg L^{-1} min. Inactivation followed a log-linear trend with a high inactivation rate, between approximately 10 and 50 mg L^{-1} min. Then, for higher doses, there is a tailing phase, with a drastic decrease of the inactivation rate.

DM, DEM, MGM1 and MGM2 were calibrated on collected data and coefficients, statistical significance and predictive performance indicators are reported in Table 4.

R^2_{adj}, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) results suggest that MGM1 is the best performing model in terms of prediction. MGM1 was formulated starting from the model presented by Geeraerd et al. (2000). This model was modified in MGM1 by

introducing PAA concentration in the inactivation rate equations, as described in paragraph 2.1.3 (Eq. 2.14-2.17), which makes MGM1 suitable for inactivation by disinfectants. The good fit of the model suggests that this hypothesis, originally formulated for thermal inactivation, could be extended to chemical disinfection, where C could work as an aggregated parameter accounting, as an example, for the presence of several reactive sites to be oxidized on bacteria cell membranes and of the presence of external polymeric substances (EPS) protecting the cell membrane (*inter alia*: Kitis, 2004, Xue et al., 2013). MGM1 was conceived to find a triphasic inactivation model, which can be used in a Eulerian model in dynamic simulations, like for the dispersion model used in this work. Inactivation rates defined so far in the literature, which can describe both lag, log-linear and tailing inactivation phases, are explicitly dependent on time or dose. Examples are the well-known Hom Model (Gyürek & Finch, 1998) or the recent inactivation rate developed by Elhalwagy et al. (2021). This kind of inactivation rates would require a Lagrangian approach, where the time spent by each simulated bacteria particle in the reactor is known. Differently, a Eulerian model needs an inactivation rate which only depends on model state variables, like the MGM1. Moreover, the integral of the MGM1 for batch conditions has an analytical solution (Eq. 2.19) which depends only on the PAA dose, being consistent with the assumption on the primary importance of the dose usually made in disinfection modeling and providing a useful simplified representation of the inactivation phenomena.

MGM2 shares most of the characteristics of MGM1, except for the definition of C decay rate, which is a first order not depending on PAA concentration. This assumption implies that the decay of all the components, which protect bacteria cells from inactivation determining a lag, depends only on time, whatever is the concentration of PAA. MGM2 can catch a first lag phase with lower inactivation rate, but lag is less sharp and the transition to the loglinear phase is smoother than for MGM1 (Figure 5b). This is the main cause of lower predictive performance with respect to MGM1. MGM2 still has the advantage of being suitable for a Eulerian model, although the analytical solution of its integral in batch boundary conditions is not available.

The DM proved lower but similar performance to MGM1, but it needs only 3 instead of 4 parameters. Moreover, all DM parameters are statistically significant, while parameter X_0 of MGM1 (as for MGM2) results as not statistically significant. The fundamental drawback of DM is that it was derived directly in its integral form to describe disinfection in batch conditions (Dominguez Henao et al., 2018b) and a derivative form of the inactivation rate is not available.

Finally, predictive performances of DEM were the lowest, since it cannot account for the presence of the initial lag. Due to its limited flexibility in describing the lag phase, performance of MGM2 were only slightly higher than DEM.

Overall, the comparison of those four inactivation models evidenced how different choices dramatically impact predicted inactivation for doses corresponding to lag and log-linear phases, which in this case study are lower than approximately 50 mg L^{-1} min. Differently, all models are approximately equal for higher doses. These considerations are evident in Figure 5b.

DM, as the model with the lowest number of parameters and the highest predictive performance (as MGM1), was used to model continuous flow disinfection in stationary conditions, according to the SFM approach (paragraph 3.4). Since DM does not have a derivative form, DEM, MGM1 and MGM2 were used and compared to model continuous flow disinfection in dynamic conditions (paragraph 3.6).

Despite data clearly suggest the general triphasic trend of the disinfection kinetics and were the basis for the calibration of models which were mostly statistically significant, variability of residuals is very relevant, as can be straightforwardly argued in Figure 5. Such variability is probably given by the intrinsic variability of effluent quality, which could result in different shapes of the kinetics and consequently different parametrizations of the triphasic kinetic model. As proven in Santoro et al. (2015) and Ahmed et al. (2019), data from single inactivation experiments on effluent samples coming from the same WWTP could be used to parametrize inactivation model, and probability distribution of parameters can be derived from parametrizations on several samples. Such approach were based on the empirical evidence that *E. coli* concentrations coming from single experiments on single samples were better aligned along calibrated inactivation models, with much lower residuals. Instead, in current work data from all inactivation experiments were pooled in a single dataset which was used for model calibration, since number of samples was not considered adequate to infer a probability distribution for parameter (n=3 in Santoro et al. (2015) and n=46 in Ahmed et al. (2019)). As additional alternative approach, a possible relationship between inactivation kinetic parameters and physical-chemical characteristics of the effluent was explored, according to the same approach studied for PAA decay kinetics (paragraphs 5.2.1.2 and 5.3.2), but no meaningful result was achieved due to inadequacy of data and, probably, of the type of wastewater parameters monitored in the study, which were in no relationship with *E. coli* inactivation mechanisms.

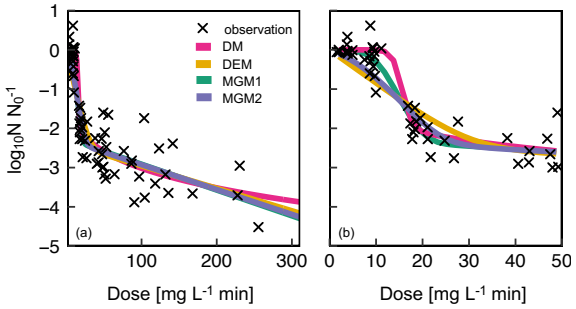

Figure 5 – Fit of experimental data by all the considered inactivation model. a) reports model fit and observations over the complete explored dose range (1-270 mg L^{-1} min). b) reports same model fit and observations between 0 and 50 mg L^{-1} min.

Table 4 – Results of inactivation models calibration.

Model	Coefficients	*p-value*	R^2_{adj}	AIC	BIC	n
Dose (DM)	k' = 1.091	<0.05	0.79	95.6	104.6	
	n = 0.221	<0.05				
	h = 15.59	<0.05				
Double Exponential (DEM)	δ = 0.996	<0.05	0.70	121.4	130.4	
	a = 0.196	<0.05				
	b = 0.013	<0.05				
Modified Geeraerd Version 1 (MGM1)	η_{max} = 0.564	<0.05	0.80	93.5	104.8	76
	X_0 = 287	0.526				
	δ = 0.994	<0.05				
	b = 0.015	<0.05				
Modified Geeraerd Version 2 (MGM2)	η_{max} = 0.346	<0.05	0.71	119.3	130.6	
	X_0 = 2.534	0.264				
	δ = 0.995	<0.05				
	b = 0.014	<0.05				

5.3.4 Prediction of PAA and E. coli residuals in stationary conditions

The use of the SFM to predict PAA and *E. coli* concentrations at the outlet of the pilot reactor was validated in experiments under stationary PAA dosage and flow rate conditions. Plots of observed against estimated PAA and *E. coli* concentrations are reported in Figure 6. Two modes of use of the SFM were compared. In the first case, PAA and *E. coli* were estimated considering PAA decay as stationary and independent from wastewater quality: k and OD calibrated on the pooled data coming from all PAA decay batch experiments were used. In the second mode, PAA decay parameters k and OD were estimated from current effluent quality parameters using the regression model described in paragraph 3.2. Results evidenced how accounting for the impact of varying wastewater quality improved predictive performance. R^2 increased from -0.22 to 0.28 for PAA concentration and from 0.86 to 0.88 for *E. coli* prediction. Despite the improvement, predictive performance of PAA residual is still globally poor, while prediction accuracy of *E. coli* remains low in higher concentration cases, as can be observed in Figure 6b. One possible explanation is that wastewater quality was not stationary throughout the pilot disinfection experiments, which is not accounted by SFM, and suggesting the importance of modeling the disinfection reactor as a dynamic system, where fluctuations in wastewater quality have to be considered when predicting disinfectant and bacteria residuals at given time. However, low predictive performance of SFM for PAA residual and the very limited improvement in predicting *E. coli* residual are also caused by partial inadequacy of data available for validation in stationary conditions. While combination of

various meaningful flow rate and PAA dosage concentration were covered (see paragraph 5.2.1.4), yet more replicates would be needed for the dataset to be more representative of effluent quality variability and, thus, to better test possible benefits of integrating wastewater quality-dependent PAA kinetic parameters in the disinfection model.

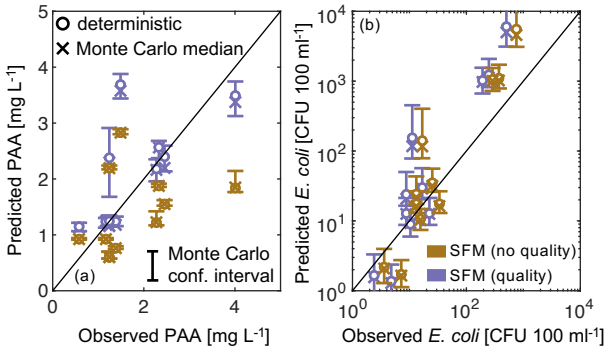

Figure 6 – Results of Segregated Flow Model (SFM) approach under stationary conditions.

5.3.5 Prediction of PAA and E. coli residuals in dynamic conditions

The model of PAA disinfection was validated on continuous flow experiments in the pilot reactor with variable effluent flow rate and quality (paragraph 2.1.5). Time trends of flow rate and turbidity, as most relevant effluent quality parameter in determining both PAA decay and initial *E. coli* concentration, are reported in Figure 7a-b. The disinfection model proved good prediction accuracy both for initial *E. coli* and PAA concentration at the outlet of the pilot reactor.

As reported in Figure 7e-f, PAA concentration was well approximated (R^2 = 0.84), but some fluctuations were not caught by the model. Results were compared to predictions of PAA residuals when neglecting the impact of effluent quality on PAA decay and assuming constant k and OD, calibrated on pooled lab experiments. As shown in Figure 7e-f, this simplification decreased prediction accuracy of PAA residual (R^2 = 0.57). Most of the worsening is due to predictions in case of higher dosage, as expected since PAA follows a first order decay rate. Real-time update of PAA decay according to effluent quality avoided underestimation of PAA residual.

Predictions of *E. coli* residuals using DEM, MGM1 and MGM2 as inactivation rates were compared and MGM2 resulted the best option in terms of predictive performance (R^2 = 0.58). As reported in Figure 7c-d, the model can reproduce *E. coli* and PAA dynamics, accounting for the impact of fluctuations in flow rate and wastewater quality. As can be seen in Figure 7c, MGM2 is the best option since it better predicts *E. coli* fluctuations both at 1 and 3 mg L^{-1}. As already argued from inactivation models in batch conditions, the choice of the inactivation rate model impacted mainly the prediction at low dosage, while predictions at higher dosage were approximately equivalent (Figure 7d). At low PAA dosage, even small variations of contact time or decay rate can significantly affect the inactivation rate, which can shift among the lag, log-linear and tailing phase.

Differently, higher dosage moves the inactivation rate in the tailing phase, with lower impact of dose variations. Importantly, *E. coli* final concentrations were predicted correctly also thanks to accurate predictions of initial *E. coli* concentration (Figure 7c-d).

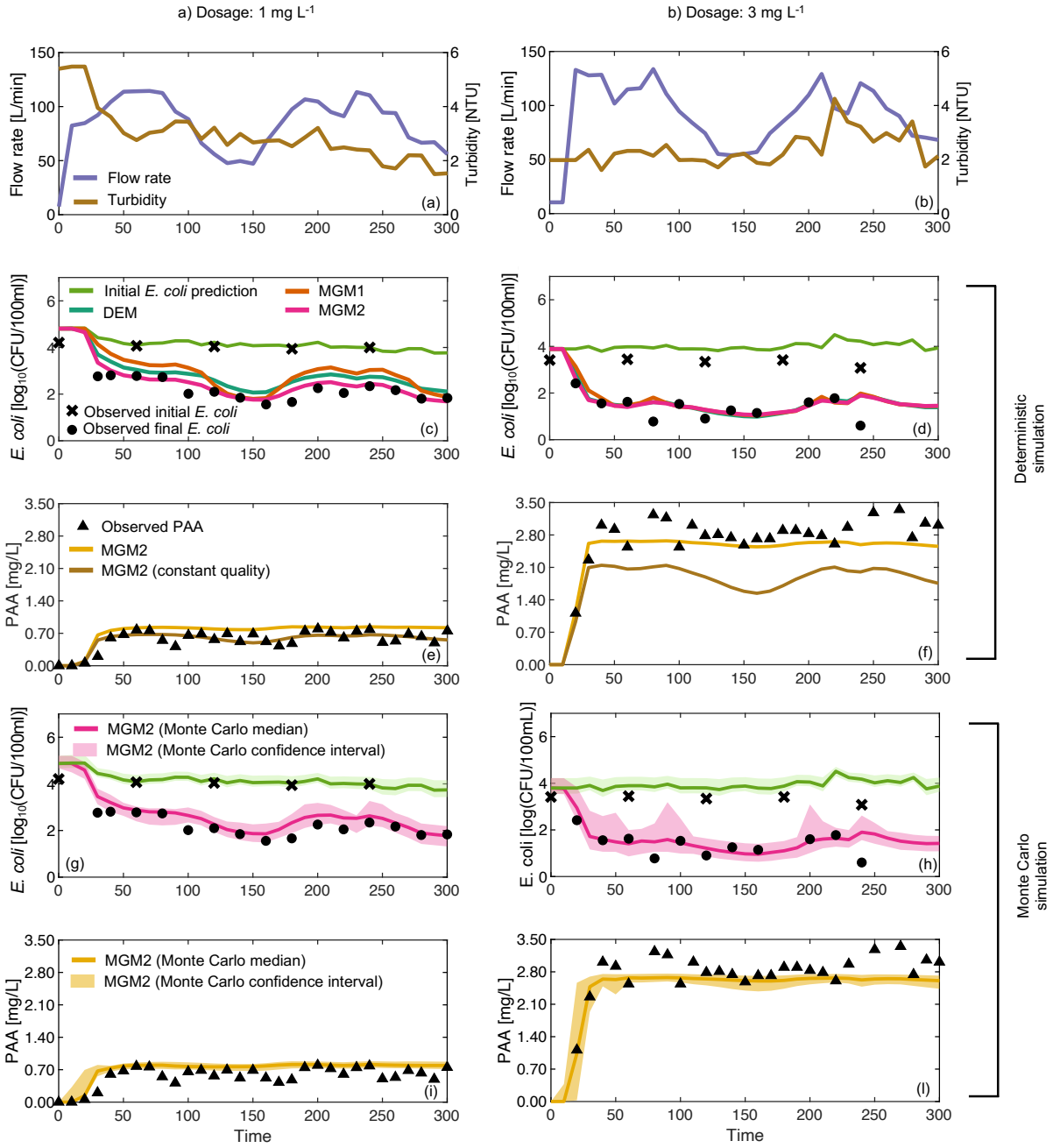

Figure 7 – Validation of the PAA disinfection model on continuous flow pilot-scale disinfection experiments. a-b) flow rate and turbidity trends during pilot experiments. c-d) deterministic validation of prediction of *E. coli* concentration at the inlet and the outlet of the pilot reactor. e-f) deterministic validation of the predictions of PAA concentration at the outlet of the reactor. g-l) validation of predictions of *E. coli* and PAA concentration with Monte Carlo simulation.

Monte Carlo simulations propagated model uncertainty onto predictions, as reported in Figure 7g-l, providing more robust estimate of disinfection outputs. Particularly, confidence interval comprises most of *E. coli* concentration data. Differently, still some fluctuations in PAA concentrations do not fall inside Monte Carlo confidence intervals.

5.4 Conclusions

In this work, a mechanistic model for continuous flow PAA disinfection was developed, calibrated and validated. The model proved good performances both in stationary and dynamic effluent flow rate and quality. The model was conceived from some well-established models, but it introduces some key innovations to improve wastewater chemical disinfection:

- Reactor hydraulics was described as two 1-dimensional dispersion model in parallel, with flow rate-varying dispersion parameters, which estimate in real-time how flow rate variations impact contact tank hydrodynamics.
- Impact of dosage and varying effluent quality on PAA decay kinetics was estimated by a linear regression model, which was used to update in real-time PAA decay parameters.
- Several *E. coli* inactivation kinetics models were compared. Different models lead to significantly different predictions at low PAA dosages, when the triphasic behavior of the kinetics is impacting.
- Two new inactivation kinetic models (MGM1 and MGM2) were proposed, which were developed from the model developed by Geeraerd et al. (2000) for thermal inactivation. Both models can be integrated in Eulerian models, like dispersion models, and can describe the triphasic behavior of inactivation by chemicals like PAA.
- The developed disinfection model was used in dynamic numerical simulations and validated on pilot-scale experiments. Successful validations highlighted the importance of modeling the disinfection process as a dynamic system, while integrating regression models to real-time update model inputs and parameters, being initial bacteria concentration, disinfectant decay and hydraulics parameters.

In light of these innovations and advantages, the developed disinfection model could be very useful for off-line optimization or on-line control of chemical disinfection, to face the always increasing need of minimization of both environmental impact and cost.

References

Antonelli, M., Turolla, A., Mezzanotte, V., & Nurizzo, C. (2013). Peracetic acid for secondary effluent disinfection: A comprehensive performance assessment. Water Science and Technology, 68(12), 2638–2644. https://doi.org/10.2166/wst.2013.542

Crittenden, J. C., Trussell, R. R., Hand, D. W., Howe, K. J., & Tchobanoglous, G. (2012).

Principles of Reactor Analysis and Mixing. MWH's Water Treatment, 287–390. https://doi.org/10.1002/9781118131473.ch6

Domínguez Henao, L., Cascio, M., Turolla, A., & Antonelli, M. (2018a). Effect of suspended solids on peracetic acid decay and bacterial inactivation kinetics: Experimental assessment and definition of predictive models. Science of the Total Environment, 643, 936–945. https://doi.org/10.1016/j.scitotenv.2018.06.219

Domínguez Henao, L., Cascio, M., Turolla, A., & Antonelli, M. (2018b). Effect of suspended solids on peracetic acid decay and bacterial inactivation kinetics: Experimental assessment and definition of predictive models. Science of the Total Environment, 643, 936–945. https://doi.org/10.1016/j.scitotenv.2018.06.219

Domínguez Henao, L., Delli Compagni, R., Turolla, A., & Antonelli, M. (2018). Influence of inorganic and organic compounds on the decay of peracetic acid in wastewater disinfection. Chemical Engineering Journal, 337(October 2017), 133–142. https://doi.org/10.1016/j.cej.2017.12.074

Du, Y., Lv, X. T., Wu, Q. Y., Zhang, D. Y., Zhou, Y. T., Peng, L., & Hu, H. Y. (2017). Formation and control of disinfection byproducts and toxicity during reclaimed water chlorination: A review. Journal of Environmental Sciences, 58, 51–63. https://doi.org/10.1016/j.jes.2017.01.013

Ducoste, J., Carlson, K., & Bellamy, W. (2001). The integrated disinfection design framework approach to reactor hydraulics characterization. Journal of Water Supply: Research and Technology - AQUA, 50(4), 245–261. https://doi.org/10.2166/aqua.2001.0021

Efron, B., & Hastie, T. (2016). Computer age statistical inference: Algorithms, evidence, and data science. In Computer Age Statistical Inference: Algorithms, Evidence, and Data Science. https://doi.org/10.1017/CBO9781316576533

Elhalwagy, M., Biabani, R., Bertanza, G., Wisdom, B., Goldman-Torres, J., McQuarrie, J., Straatman, A., & Santoro, D. (2021). Mechanistic modeling of peracetic acid wastewater disinfection using computational fluid dynamics: Integrating solids settling with microbial inactivation kinetics. Water Research, 201(June), 117355. https://doi.org/10.1016/j.watres.2021.117355

Foschi, J., Turolla, A., & Antonelli, M. (2021). Soft sensor predictor of E. coli concentration based on conventional monitoring parameters for wastewater disinfection control. Water Research,

191, 116806. https://doi.org/10.1016/j.watres.2021.116806

Furman, L., Leclerc, J. P., & Stegowski, Z. (2005). Tracer investigation of a packed column under variable flow. Chemical Engineering Science, 60(11), 3043–3048. https://doi.org/10.1016/j.ces.2004.12.024

Geeraerd, A. H., Herremans, C. H., & Van Impe, J. F. (2000). Structural model requirements to describe microbial inactivation during a mild heat treatment. International Journal of Food Microbiology, 59(3), 185–209. https://doi.org/10.1016/S0168-1605(00)00362-7

Gorzalski, A. S., Harrington, G. W., & Coronell, O. (2018). Modeling Water Treatment Reactor Hydraulics Using Reactor Networks. Journal - American Water Works Association, 110(8), 13–29. https://doi.org/10.1002/awwa.1071

Gyürek, L. L., & Frinch, G. R. (1998). MODELING WATER TREATMENT CHEMICAL DISINFECTION KINETICS By Lyndon L. Gyiirek 1 and Gordon R. Finch/ Member, ASCE. Journal of Environmental Engineering, 124(SEPTEMBER), 783–793.

Hastie, T. et. all. (2009). Springer Series in Statistics The Elements of Statistical Learning. The Mathematical Intelligencer, 27(2), 83–85. https://doi.org/10.1007/b94608

How, Z. T., Kristiana, I., Busetti, F., Linge, K. L., & Joll, C. A. (2017). Organic chloramines in chlorine-based disinfected water systems: A critical review. Journal of Environmental Sciences (China), 58, 2–18. https://doi.org/10.1016/j.jes.2017.05.025

John C. Crittenden, R. Rhodes Trussell, David W. Hand, K. J. H. and G. T. (2017). MHW´s Water Treatment Principles and Design.

Kadoya, S. S., Nishimura, O., Kato, H., & Sano, D. (2020). Regularized regression analysis for the prediction of virus inactivation efficiency by chloramine disinfection. Environmental Science: Water Research and Technology, 6(12), 3341–3350. https://doi.org/10.1039/d0ew00539h

Kadoya, S. suke, Nishimura, O., Kato, H., & Sano, D. (2021). Predictive water virology using regularized regression analyses for projecting virus inactivation efficiency in ozone disinfection. Water Research X, 11, 100093. https://doi.org/10.1016/j.wroa.2021.100093

Kitis, M. (2004). Disinfection of wastewater with peracetic acid: A review. Environment International, 30(1), 47–55. https://doi.org/10.1016/S0160-4120(03)00147-8

Levenspiel, O., 1999. Chemical reaction engineering, Third. ed. John Wiley & Sons, Toronto.

Manoli, K., Sarathy, S., Maffettone, R., & Santoro, D. (2019). Detailed modeling and advanced

control for chemical disinfection of secondary effluent wastewater by peracetic acid. Water Research, 153, 251–262. https://doi.org/10.1016/j.watres.2019.01.022

McFadden, M., Loconsole, J., Schockling, A. J., Nerenberg, R., & Pavissich, J. P. (2017). Comparing peracetic acid and hypochlorite for disinfection of combined sewer overflows: Effects of suspended-solids and pH. Science of the Total Environment, 599–600, 533–539. https://doi.org/10.1016/j.scitotenv.2017.04.179

Neumann, M. B., von Gunten, U., & Gujer, W. (2007). Uncertainty in prediction of disinfection performance. Water Research, 41(11), 2371–2378. https://doi.org/10.1016/j.watres.2007.02.022

Newhart, K. B., Goldman-Torres, J. E., Freedman, D. E., Wisdom, K. B., Hering, A. S., & Cath, T. Y. (2021). Prediction of Peracetic Acid Disinfection Performance for Secondary Municipal Wastewater Treatment Using Artificial Neural Networks. ACS ES&T Water, 1(2), 328–338. https://doi.org/10.1021/acsestwater.0c00095

Rossi, S., Antonelli, M., Mezzanotte, V., & Nurizzo, C. (2007). Peracetic Acid Disinfection: A Feasible Alternative to Wastewater Chlorination. Water Environment Research, 79(4), 341–350. https://doi.org/10.2175/106143006x101953

Santoro, D., Bartrand, T. A., Greene, D. J., Farouk, B., Haas, C. N., Notarnicola, M., & Liberti, L. (2005). Use of CFD for wastewater disinfection process analysis: E.coli inactivation with peroxyacetic Acid (PAA). International Journal of Chemical Reactor Engineering, 3(August 2016). https://doi.org/10.2202/1542-6580.1283

Santoro, D., Crapulli, F., Raisee, M., Raspa, G., & Haas, C. N. (2015). Nondeterministic Computational Fluid Dynamics Modeling of Escherichia coli Inactivation by Peracetic Acid in Municipal Wastewater Contact Tanks. Environmental Science and Technology, 49(12), 7265–7275. https://doi.org/10.1021/es5059742

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

USEPA. (2006). National Primary Drinking Water Regulations: Long term 2 Enhanced Surface Water Treatment Rule (40 CFR Parts 9, 141, and 142). Federal Register, 71(3), 654–786.

Wang, Y. H., Wu, Y. H., Du, Y., Li, Q., Cong, Y., Huo, Z. Y., Chen, Z., Yang, H. W., Liu, S. M., & Hu, H. Y. (2019). Quantifying chlorine-reactive substances to establish a chlorine decay

model of reclaimed water using chemical chlorine demands. Chemical Engineering Journal, 356(August 2018), 791–798. https://doi.org/10.1016/j.cej.2018.09.091

Wei, W., Haas, C. N., & Farouk, B. (2020). Development of a CFD-Based Artificial Neural Network Metamodel in a Wastewater Disinfection Process with Peracetic Acid. Journal of Environmental Engineering, 146(12), 04020140. https://doi.org/10.1061/(asce)ee.1943-7870.0001822

White, G. C. (2010). White's handbook of Chlorination and Alteratie Desinfectants. In Wiley, Hoboken, N.J. https://doi.org/10.1002/9780470561331.ch7

Xue, Z., Hessler, C. M., Panmanee, W., Hassett, D. J., & Seo, Y. (2013). Pseudomonas aeruginosa inactivation mechanism is affected by capsular extracellular polymeric substances reactivity with chlorine and monochloramine. FEMS Microbiology Ecology, 83(1), 101–111. https://doi.org/10.1111/j.1574-6941.2012.01453.x

Yang, X., Guo, W., & Lee, W. (2013). Formation of disinfection byproducts upon chlorine dioxide preoxidation followed by chlorination or chloramination of natural organic matter. Chemosphere, 91(11), 1477–1485. https://doi.org/10.1016/j.chemosphere.2012.12.014

Zheng, M., & Mackley, M. (2008). The axial dispersion performance of an oscillatory flow meso-reactor with relevance to continuous flow operation. Chemical Engineering Science, 63(7), 1788–1799. https://doi.org/10.1016/j.ces.2007.12.020

5.5 Supplementary material

5.5.1 Additional results on hydraulic model calibration

Table S1 – Results of the calibration of dispersion models SDM and PDM over tracer experiments

Flow rate	HRT	SDM		PDM				
		D	R^2_{adj}	D_1	D_2	α	β	R^2_{adj}
40	56	2.15	0.462	0.56	0.93	0.51	0.40	0.861
66	34	1.19	0.736	0.50	0.90	0.65	0.55	0.988
80	29	0.73	0.748	0.39	2.92	0.81	0.74	0.981
90	25	0.50	0.699	0.36	4.02	0.88	0.80	0.972
130	17	0.29	0.946	0.24	10.88	0.87	0.89	0.961

5.5.2 Additional results on peracetic acid decay batch experiments

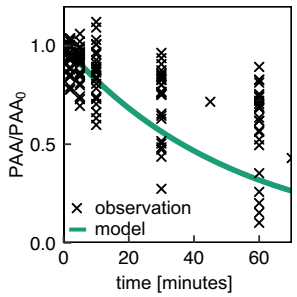

Figure S1 – Pooled data from lab scale batch PAA decay experiments and Haas and Finch model fit.

5.5.3 Additional details on model uncertainty

Model uncertainty was assessed by *non-parametric bootstrap* (Efron & Hastie, 2016). Parameters of each model were repeatedly tuned on n bootstrap samples with replacement independently drawn from the original data, resulting in n different estimates $\hat{p}_i$ of each parameter p_i. Probability distribution and confidence interval of p_i were then approximated by the observed distribution of $\hat{p}_i$.

Model uncertainty was propagated on the predictions of the continuous flow disinfection model (paragraph 2.1.6) with Monte Carlo simulation of size n. Confidence intervals were estimated as delimited by 5th and 95th percentiles of model outputs simulation, while point estimates were computed as the median of model outputs simulations, according to the *bragging* approach (Bühlmann, 2003). Results from the combination of bootstrap resampling and Monte Carlo simulations were compared with deterministic simulations, where model parameters were tuned without bootstrap resampling.

5.5.4 Additional details on calibration of the initial E. coli concentration model

E. coli concentration observed in the 19 undisinfected samples of the WWTP effluent were highly variable, ranging between 1,300 and 730,000 CFU 100 ml^{-1}. Since such variability can drastically impact disinfection efficiency, a regression model was calibrated on collected data to predict *E. coli* concentration using monitored effluent quality parameters as predictors, according to a "soft-sensor" approach which was already explored in similar applications (Foschi et al., 2021). As a result of LASSO regression and LOOCV procedure, 3 variables were selected as predictors (Figure 4a), being pH, TRB and NO_3^-, whose coefficients are reported in Table 3. As can be argued by the absolute value of coefficients, TRB is the most important predictor. As supposed in Foschi et

al. (2021), the increase in TRB could determine a higher contribution of the particle-associated bacteria fraction in the effluent, due to temporary decrease of upstream treatments efficiencies.

Chapter 6: Model predictive control of wastewater disinfection by peracetic acid

Abstract

Wastewater disinfection is designed to reduce, under given regulatory limit, the concentration of fecal indicators microorganisms, which is not directly measurable online. Then, model-based control is needed for real-time optimization of disinfectant dosage. In this work, feedforward model predictive control (MPC) of peracetic acid (PAA) disinfection of a real tertiary effluent from a municipal wastewater treatment plant was conceived and its effectiveness was tested at pilot scale, considering *E. coli* as fecal indicator. Previously developed PAA disinfection model was integrated in the MPC scheme, which repeatedly optimizes PAA dosage at each time instant, based on model predictions over a finite time horizon and measurements of effluent flow rate and quality and residual PAA concentration. MPC accounts for time-varying dynamic nature of the process, considering how changes in flow rate affect reactor hydraulics, while fluctuations in wastewater quality impact PAA decay. Moreover, MPC can handle the highly nonlinear triphasic disinfection kinetics of *E. coli*. Pilot scale control experiments proved the effectiveness of MPC control in meeting *E. coli* discharge limits of 10 and 100 CFU/100 mL. Comparison with traditional *flow-paced* control of disinfection, which maintains a constant dosage, revealed that MPC control, besides compensating time-varying operating conditions by optimizing dosage in real-time, can potentially save from 30% to 85% of disinfectant. Results suggest that MPC control is particularly important in case of the strictest among studied discharge limits, i.e. 10 and 100 CFU/100 mL, where it potentially provides highest savings and significantly reduces average PAA dosage, limiting the occurrence of potentially ecotoxic conditions at the discharge. MPC control could be then an effective option for wastewater reuse applications, which demands for high disinfection efficiency, balancing process efficiency, cost and sustainability.

This chapter will be submitted soon for publication.

6.1 Introduction

Raw sewage, sewer overflows and inadequately treated wastewater treatment plant (WWTP) effluents are main causes of fecal contamination of surface waters (*inter alia*: Naidoo & Olaniran, 2013), impacting human health by impairing sources of drinking and supply water, and bathing areas (*inter alia*: Baum et al., 2013). Increasing global water scarcity is promoting the spread of direct and indirect potable and agricultural reuse practices, which can reduce human water footprint at the price of implementation of measures to limit potential health risk from exposure to enteric pathogens (Rodriguez et al., 2009; Angelakis et al., 2018). Given this scenario, in last decades and still today multilateral institutions, environmental agencies and national governments published regulations and guidelines defining ever more strict limits on fecal indicators (FI) for WWTPs (Lyu et al., 2016, Lavrnić et al., 2017, Ritter, 2021, Radcliffe, 2022). Not only compliance with low FI limits is challenging, but violations could have severe impacts on human health (Onyango et al., 2015, Rock et al., 2019). Control of disinfection reactors in WWTPs is then of primary importance since it is devoted to reducing FI concentration in treated effluent before discharge or reuse.

In chemical disinfection, which is the most widespread type, concentration of disinfectant and contact time drive inactivation efficiency of the system. Main process disturbances are flow rate, affecting hydraulic residence time in the reactor, initial concentration of FI, determining required reduction to reach given discharge limit, and wastewater chemical and physical-chemical parameters, which affect disinfectant decay (*inter alia*: Domínguez Henao, Delli Compagni, et al., 2018, Domínguez Henao, Cascio, et al., 2018). Inactivation kinetics determines the log-reduction of FI at given initial FI concentration, disinfectant concentration and contact time, and its shape depends on both the type of disinfectant and FI (*inter alia*: Balachandran et al., 2021). Disinfectant concentration has to be controlled in order to compensate disturbances and keep outlet FI concentration under the compliance limit, according to the inactivation kinetics. Main references for disinfection dosage design are the Surface Water Treatment Rule (USEPA, 2006) and the Integrated Disinfection Design Framework (IDDF) (Ducoste et al., 2001), but they both provide a modeling framework only for off-line optimization of dosage. On-line control in wastewater treatment was extensively studied (Santín et al., 2015, Han et al., 2014, Zeng & Liu, 2015), but applications exist mainly for biological removal of organics and nutrients and chemical oxidation and coagulation processes, neglecting disinfection. More recently, machine learning methods and fuzzy logic control were applied to chemical disinfection *flow-paced* control in order to compensate the disturbances from the presence of ammonia and nitrite in the effluent (Khawaga et al., 2019). Manoli et al. (2019) summarized recent studies on on-line control of disinfectant concentration,

which are mainly based on feedforward and feedback *flow-paced* control of disinfectant, maintaining constant initial concentration of disinfectant (*inter alia*: Kobylinski et al., 2014, Dieu et al., 1995, Block et al., 2015). It was recently demonstrated that a shift from *flow-paced* to *dose-paced* control is fundamental for reliable control of disinfection (Domínguez Henao, Cascio, et al., 2018, Manoli et al., 2019). Since inactivation is determined by both concentration of disinfectant and contact time, the *dose*, defined as the integral of PAA concentration up to current contact time, which synthesize the effect of both the mentioned operating parameters, should be controlled to maintain compliance on outlet FI concentration.

Despite the advances described so far, still feedback, and possibly *dose-based*, control of disinfection which considers the reactor as a dynamic system was not developed and control reliability has not been tested under relevant real-time variations of effluent quality and flow rate. This work aims at defining and testing feedforward and feedback model predictive control (MPC, Camacho & Bordons, 1999, Eaton & Rawlings, 1992) of wastewater disinfection, with application to peracetic acid (PAA) disinfection of tertiary effluent, assuming *E. coli* as FI. PAA inactivation kinetics is highly nonlinear, while PAA decay kinetics and hydraulic behavior of the contact tank change with effluent quality and flow rate, respectively. MPC was then chosen as control approach since it can effectively handle nonlinear and time varying processes, by using nonlinear models with time-varying parameters. Moreover, disinfection performance at each instant is given by the current concentration of FI at the outlet of the reactor, which is not measurable in real-time. FI concentration can only be estimated by a model and model-based optimization is then needed to calculate disinfectant dosage to meet the target setpoint. Previously developed PAA disinfection model was integrated in a standard Receding Horizon MPC control scheme. The model is the combination of two parallel 1-dimensional *advection-dispersion-reaction* models, which considers both the effect of PAA concentration and contact time for FI inactivation prediction, thus respecting assumptions of *dose-based* control. The control algorithm was deployed on a pilot scale reactor, fed by real undisinfected tertiary effluent. Pilot-scale control experiments were carried out to test MPC effectiveness in meeting the strictest regulation limits of the new European regulation (The European Parliament and the Council of the European Union, 2020) for agricultural water reuse. Real undisinfected tertiary effluent was fed the pilot disinfection reactor, with real-time changing wastewater quality, and flow rate patterns were artificially generated to realize typical average hydraulic residence time of real scale disinfection reactors. Finally, benefits of real time control of PAA disinfection were assessed by comparing PAA consumption in case of MPC control with off-line optimization scenarios.

6.2 Materials and methods

6.2.1 PAA disinfection model

Previously developed and validated PAA disinfection model (Foschi et al. 2022) was used in this work for MPC control of pilot scale PAA disinfection. The pilot reactor was described as two parallel 1-dimensional *advection-dispersion-reaction* model, also known as *dispersed plug-flow* or *dispersion* model (Levenspiel, 1999, Crittenden et al., 2017). Total flow rate and cross-sectional area of the reactor are split between the two dispersion models according to two parameters α and β, respectively, with $\alpha, \beta \in [0,1]$. The model assumes that *E. coli* bacteria population is divided into *free-swimming* (FS) and *particle-associated* (PA) bacteria. The first sub-population inactivation kinetics is characterized by a first *lag* phase, with lower inactivation rate, followed by a log-linear phase with higher inactivation rate. The lag phase is assumed to be caused by the presence of a group of components (X) outside and inside bacteria cells which protect from the disinfectant action. X decays as a first order kinetics. The second sub-population follows a second order inactivation kinetics with much slower rate, determining a *tailing* phase in inactivation kinetics of total *E. coli*. The proportions of FS and PA *E. coli* are given by γ and $1-\gamma$, respectively, with $\gamma \in [0,1]$. PAA decay follows first order decay rate, whose constant depends on PAA dosage and disinfected effluent quality $\bar{q} = [PAA_0, TRB, ABS_{254}, NO_3^-, TN]$, where PAA_0 (mg L^{-1}) is PAA dosage, TRB (NTU) is turbidity, ABS_{254} is absorbance at 254 nm, NO_3^- (mg L^{-1}) is nitrate concentration and TN (mg L^{-1}) is total nitrogen concentration. PAA concentration (mg L^{-1}), *E. coli* concentration (N, CFU/100mL) and X (-) are state variables of the dynamic system defined by the dispersion model, described by the following system of PDEs:

$$\frac{dPAA_i}{dt} = -v\frac{dPAA_i}{dx} + D_i(Q_i)\frac{d^2PAA_i}{dx^2} - k(\bar{q})PAA_i \quad \text{(Eq. 2.1)}$$

$$\frac{dX_i}{dt} = -v\frac{dX_i}{dx} + D_i(Q_i)\frac{d^2X_i}{dx^2} - \eta_{max}X_i \quad \text{(Eq. 2.2)}$$

$$\frac{dN_{fs,i}}{dt} = -v\frac{dN_{fs,i}}{dx} + D_i(Q_i)\frac{d^2N_{fs,i}}{dx^2} - \eta_{max}N_{fs,i}PAA_i\frac{1}{1+X_i} \quad \text{(Eq. 2.3)}$$

$$\frac{dN_{pa,i}}{dt} = -v\frac{dN_{pa,i}}{dx} + D_i(Q_i)\frac{d^2N_{pa,i}}{dx^2} - \beta N_{pa,i}PAA_i \quad \text{(Eq. 2.4)}$$

where $i \in \{1,2\}$ identifies the first and the second dispersed plug-flow reactors which constitute the overall model. As reported in the equations, dispersion parameter D (m^2 min^{-1}) depends on flow rate (m^3 min^{-1}), since different flow rate regimes impact reactor hydrodynamics (Foschi et al., 2022). Concentration of *E. coli* at the outlet of the reactor is given by:

$$N(x=L) = \alpha\left(N_{fs,1}(x=L) + N_{fs,2}(x=L)\right) + (1-\alpha)\left(N_{pa,1}(x=L) + N_{pa,2}(x=L)\right) \quad \text{(Eq. 2.5)}$$

where L is the length of the contact tank.

Open-flow boundary conditions were assumed (Crittenden et al., 2017). Initial and boundary conditions for the system of PDEs described in Eqs. 2.1-4 depend on PAA dosage PAA_0 and concentration of *E. coli* in the undisinfected effluent N_0. PAA dosage is reduced by an amount equal to the estimated oxidative demand of the effluent OD (mg L^{-1}), which is a positively correlated linear function of PAA_0 (Foschi et al., 2022).

6.2.2 MPC control of PAA disinfection

MPC control of PAA disinfection at pilot scale consisted of the solution at each time instant of the open-loop optimization of PAA dosage over a finite horizon. Discrete time is assumed, with time interval (Δt) equal to 10 minutes. The optimization problem to be solved at each time t is defined as:

$$\min_{PAA_0} J(Z_t, PAA_0) \quad \text{(Eq. 2.6)}$$

subject to:

$$J_t = \sum_{i=t}^{t+H} (log_{10}y_i - \vartheta)^{+T} Q (log_{10}y_i - \vartheta)^{+} + {u_i}^T R u_i \quad \text{(Eq. 2.7).}$$

$$Z_{t+1} = f(Z_t, u_t, w_t) \quad \text{(Eq. 2.8)}$$

$$y_t = N_t = N_{fs,t} + N_{pa,t} \quad \text{(Eq. 2.9)}$$

$$u_t = PAA_0 \quad \text{(Eq. 2.10)}$$

$$w_t = [N_{0_{obs}}, Q_{obs}, TRB_{obs}, {NO_3^-}_{obs}, TN_{obs}]^T \quad \text{(Eq. 2.11)}$$

with:

$$t \in [t, t+H] \quad \text{(Eq. 2.12)}$$

where J is the objective function to be minimized, Z is the vector of state variables PAA, X, N_{fs} and N_{pa}, the function $f(\cdot)$ is the result of the discretization of the system of PDEs described by Eqs. 2.1-4. w_t is the vector of disturbances, which are initial *E. coli* concentration $N_{0,t}$, flow rate Q_t and effluent quality parameters $TRB_t, {NO_3^-}_t, TN_t$. According to a feedforward compensation scheme, they are all assumed to be constant and equal to measured disturbances (marked by subscript "obs" in Eq. 2.11) at time t over the whole control horizon, until t+H. Since $N_{0,t}$ is not measurable on-line, a soft sensor approach was used to virtually monitor initial *E. coli* concentration (Foschi et al., 2021). As defined in Eq. 2.10, dosage from t to $t+H$ is assumed to be constant and equal to PAA_0. Thus, at each time t, the optimization problem has one decision variable, and the solution is the

optimal dosage $PAA^*_{0,t}$. As defined in Eq. 2.6, $PAA^*_{0,t}$ minimizes the standard quadratic convex objective function J, defined as the sum of squared deviation of the logarithm of N from the desired setpoint ϑ and the squared sum of the control u. Optimization is performed by Sequential Quadratic Programming (SQP), an iterative method for nonlinear optimization. MPC algorithm implies that at time t, once optimal control $PAA^*_{0,t}$ is computed, it is applied to the system only at time instant t. Then, the controlled system, i.e. the pilot disinfection reactor, evolves: at $t+1$ the new state Z_{t+1} of the system is observed and the optimization of the dosage is repeated. According to this scheme, the repeated open-loop optimization behaves as a closed-loop controller.

At each time t, only PAA_t is practically measurable on-line, since *E. coli* concentration requires culture-based methods, while X is an aggregated parameter which is introduced only for modeling need (see paragraph 2.3) and cannot be measured. Extended Kalman Filter (EKF) is then used for estimation of the whole state Z_t. At each time t, PAA_t was measured and the *a posteriori* state estimate $\hat{Z}_t$ is computed and used as initial condition for predictions from $t+1$ to $t+H$.

6.2.3 Experimental test of effectiveness of PAA disinfection MPC control

MPC control of PAA disinfection was implemented in a pilot disinfection reactor fed with real undisinfected tertiary effluent from a full-scale WWTP and thus at time varying effluent quality characteristics. Two 3-hours long experiments were carried out, where MPC was used to control effluent *E. coli* concentration to meet two setpoints, being 10 and 100 CFU/100 mL, respectively. In both experiments, flow rate varied according to fluctuating patterns (40 to 140 L min^{-1}, with HRT ranging between 17 and 55 min). PAA concentration at the outlet of the pilot reactor was measured each 10 minutes. PAA measurements were used both as input of MPC control and as validation of the PAA model one-step-ahead predictions during control experiments. *E. coli* concentration at the inlet and the outlet of the reactor was measured each 60 and 20 minutes, respectively. Measurements of *E. coli* concentration at the inlet were used to validate soft sensor estimates (see paragraph 2.1), while measurements at the outlet were used to test the effectiveness of MPC control in maintaining compliance with the discharge limit. Turbidity, conductivity and pH at the inlet of the reactor were measured each 10 minutes.

6.2.4 Description of the pilot plant and MPC control deployment

The experimental activity was set up in the Peschiera Borromeo municipal WWTP (440.000 PE), which includes a primary sedimentation stage, followed by a suspended biomass nitrification/oxidation and phosphorous removal by aluminium chloride. The stream quality is then refined by a double stage biofilter (BIOFOR®, Suez) performing nitrification, denitrification and

suspended solids filtration. The pilot scale disinfection plant was a 2.2 m^3 open channel *chicane* type reactor, whose horizontal and vertical sections are detailed in Figure 1. The plant inlet was equipped with an on-line propeller flow meter (Digiten FL-1608, China), a motor valve (Burkert 3285, Germany) and a peristaltic dosing pump (SEKO, USA) for PAA dosage. All devices were connected to an Arduino UNO programmable board for flow rate readings, control of valve opening and dosing pump flow rate. The contact tank was fed on the undisinfected tertiary effluent of the WWTP by a centrifugal pump (DAB, USA) withdrawing from the inlet of the full-scale disinfection contact tank, resulting in real-time varying effluent quality of the effluent entering the pilot reactor.

During control experiments, varying flow rate patterns were generated by means of the motor valve. Each 10 minutes, the Arduino UNO board received readings of flow rate and turbidity at the inlet of the reactor and PAA concentration at the outlet. While flow rate data were automatically sent by the flow meter, turbidity and PAA were measured manually (according to methods described in paragraph 2.6) and data were sent to the Arduino board by manual entry. The Arduino board sent recorded data to a PC, which performed state estimation and solve the optimization problem for MPC control, as it is defined in paragraph 2.2. Calculations were done in Mathworks Matlab R2021b. In details, SQP for objective function minimization was performed with the *fmincon* function from the Matlab Optimization Toolbox and the system of PDEs of the PAA disinfection model (see paragraph 2.1) was discretized and numerically solved by function *pdepe*, from the Matlab Partial Differential Equation Toolbox.

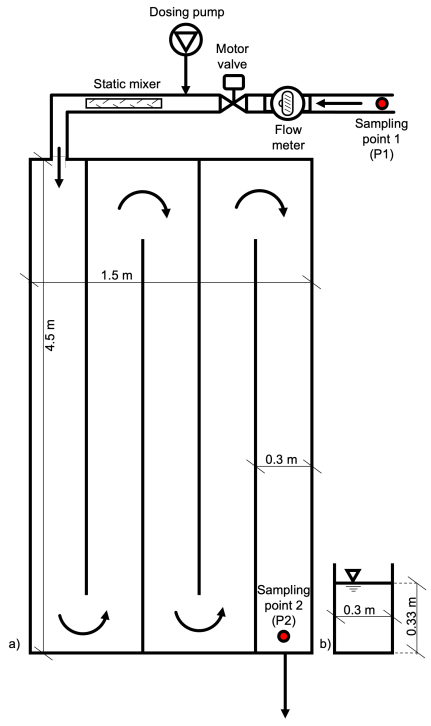

Figure 2 – Scheme of the disinfection pilot plant (not in scale).

6.2.5 Chemical and microbiological analysis

PAA residual concentration was measured by the DPD method according to Dominguez-Henao et al. (2018). Turbidity and UV254 were respectively measured by portable turbidimeter (VELP Scientifica) and 1-cm optical path laboratory spectrophotometer (Hach DR6000). pH and conductivity were measured with a multiparameter laboratory probe (HACH HQ40d).

E. coli were enumerated by membrane filtration method according to Standard Methods (Section 9222, APHA/AWWA/WEF, 2012), using 0.45- μm pore size cellulose nitrate membranes (Whatman) and chromogenic agar (Microinstant® Chromogenic Coliforms Agar, Scharlau) as culture medium. Inoculated plates were incubated at 37 °C for 24 h. *E. coli* were expressed as CFU in 100 mL volume.

6.3 Results

Effectiveness and benefits of MPC control of PAA disinfection were assessed and discussed considering the four discharge limits on *E. coli* concentration defined by the European Commission for the direct water reuse in agriculture (The European Parliament and the Council of the European Union, 2020), which are 10, 100, 1000 and 10,000 CFU/100mL. In fact, this regulation differentiates discharge limits according to the water reuse scenario, with stricter limits the more probable is the exposure to pathogens. This regulation was chosen as reference since it covers a wide range of concentration limits, equal or close to many other worldwide, and it is likely to be significantly impacting water reuse in many countries.

In paragraph 3.1, the effectiveness of MPC control in maintaining compliance with the two strictest among the considered limits, i.e. 10 and 100 CFU/100 mL, was tested in two pilot disinfection experiments.

In paragraph 3.2, MPC control for all discharge limits of the EU regulation was simulated, assuming effluent quality and flow rate recorded during the first experiment as reference scenario. Benefits and limits of MPC control were discussed in comparison with two benchmark dosage strategies (BDS). In both BDS, PAA dosage was maintained constant, assuming *flow-paced* control. PAA dosage was set as the minimum PAA concentration to meet the selected discharge limit, using the same disinfection model as for MPC control but assuming stationary operating conditions. PAA decay parameters k and OD were assumed constant, as calibrated in a previous work (Foschi et al. 2022) on pooled data collected in lab scale experiments performed on samples at different quality of the effluent under study. The two BDS differed for the assumptions made for flow rate and initial *E. coli* concentration. In the first BDS (BSD1), flow rate was assumed to be equal to the average

observed flow rate during the considered scenario (experiment 1) and initial *E. coli* concentration was assumed as equal to the average observed values for the tertiary effluent under study, previously published in Foschi et al. (2022). Differently, in second BDS (BDS2), 95th percentile was assumed instead of the average for both flow rate and *E. coli* concentration. BDS1 and BDS2 were defined as likely representations of traditional approach in designing disinfection dosage, assuming site-specific but stationary operating conditions, as in IDDF. Since it assumed average operating conditions, BDS1 is representative of a risk neutral attitude, while BDS2, which assumed the 95th percentile, is representative of a conservative approach.

6.3.1 Effectiveness of MPC control of PAA disinfection

Both experiments for testing the effectiveness of MPC control of PAA disinfection started with null concentration of PAA throughout the whole pilot reactor, with dosage starting right at the beginning of the experiment. Results reported in Figure 2 cover then MPC functioning since the initial transient phase from "undisinfected" conditions. Flow rate patterns during experiments are reported in Figures 2.a-b. Turbidity measurements are reported too, since such parameter is the most relevant predictor of both initial *E. coli* concentration and PAA decay kinetics parameters k and OD, according to the used PAA disinfection model (see paragraph 2.3.1). Figures 2.c-d report observations, soft sensor estimates and one-step-ahead predictions of concentration of initial and final concentration of *E. coli* in the pilot reactor. Figures 2.e-f report PAA dosage and observations, and one-step-ahead predictions of PAA concentration at the outlet of the reactor.

MPC control performed well during the first experiment (Figure 2.a, 2.c and 2.e), when the setpoint concentration of *E. coli* at discharge was 10 CFU/100mL. As reported in Figure 2.c, *E. coli* concentration stabilized very close to the setpoint, after the initial transient phase. MPC control was then able to compensate variations in effluent quality and flow rate, maintaining process efficiency very close to the target. During the second experiment, when a discharge limit of 100 CFU/100mL was assumed, MPC was effective only until minute 140 and 160, when two violations of the limit were recorded, with concentrations significantly higher than the setpoint. The reason of the violations is a failure of the PAA dosing pump, which stopped when PAA dosage reached its minimum. As reported in Figure 2.d, the PAA disinfection model can correctly predict the increase in *E. coli* concentration and the violation if the temporary dosage interruption is considered. In both experiments, *E. coli* soft sensor (see paragraph 2.2) overestimated initial *E. coli* concentration, which resulted in a slight overdosage and a final concentration lower than the setpoint in many cases. *E. coli* soft sensor is a data-driven regression model predicting *E. coli* concentration from effluent quality, with a predominant role of turbidity. Probably, in these experiments the effluent

under study changed too significantly with respect to conditions observed during soft sensor calibration so that the turbidity did not behave as a good predictor.

Figures 2.e-f show how MPC control tuned PAA dosage to maintain target *E. coli* concentration, compensating variations in flow rate and turbidity, with PAA dosage generally increasing with these parameters. In fact, higher flow rate implies lower contact time, while higher turbidity is correlated to higher PAA decay and initial *E. coli* concentrations. The PAA disinfection model provided good one-step-ahead predictions of PAA concentration at the outlet of the pilot reactor.

Previous research proved the effectiveness of a disinfection control algorithm which incorporate the dose concept in maintaining *E. coli* concentration under a pre-defined setpoint. In Manoli et al. (2019) outlet *E. coli* concentration in PAA disinfection experiments shew some fluctuations which could not be compensated by dosage control, but it was always compliant with the setpoint, differently from experiments presented in these paragraphs, where concentration sometimes approached and slightly passed the compliance limit, like recorded by previous literature for *flow-paced* control experiments of PAA disinfection (Block et al., 2015). Similarly, successful *dose-paced* control of *Cryptosporidium* by ozone disinfection was achieved in Dieu et al. (2012), but still final concentration of target microorganism during experiments was not stable around the setpoint. Possibly, this could be due to the impact of disturbances which are not incorporated in models used for control or which are purely stochastic. Next necessary step will be then the shift to a Robust MPC control scheme, which incorporates a more conservative attitude in the control algorithm and achieves compliance admitting a pre-defined and acceptable probability of non-compliance (Camacho & Bordons, 1999).

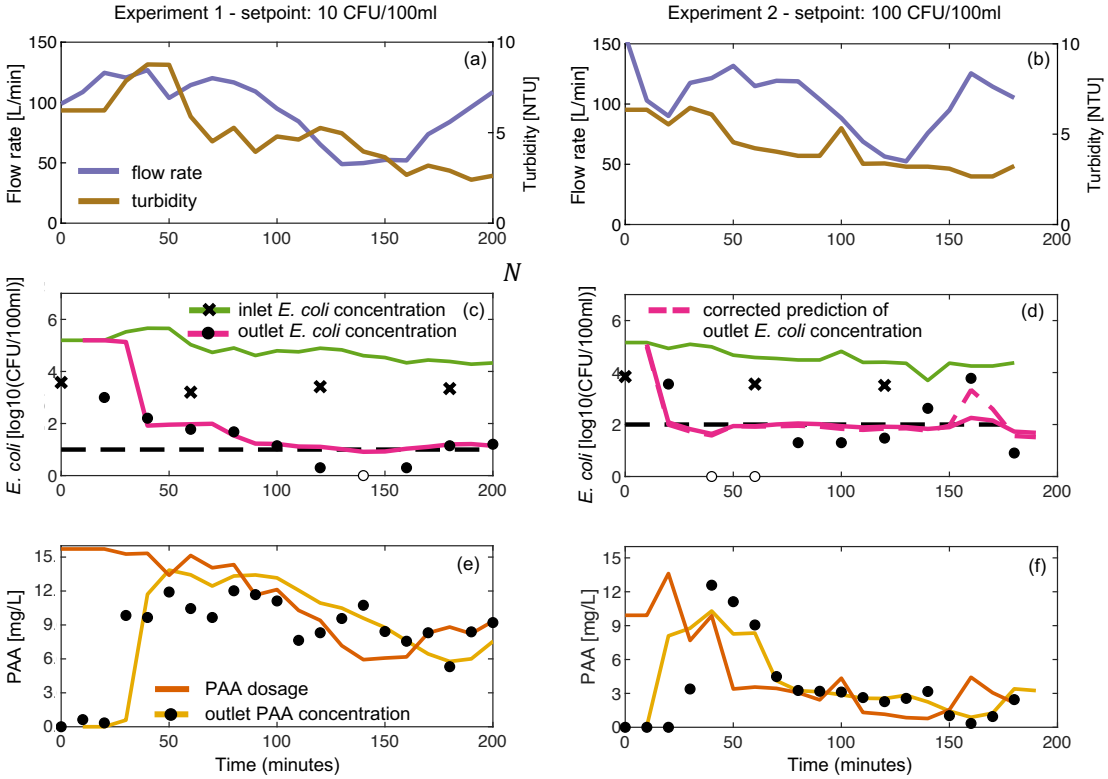

Figure 2 – Results from MPC experiments at pilot scale. a-b) flow rate and turbidity trends during pilot experiments. c-d) observed and predicted inlet and outlet concentration of *E. coli*. e-f) PAA dosage and observed and predicted concentration of PAA at the outlet of the pilot.

6.3.2 Simulation of MPC control to meet water reuse limits

Three different dosage strategies, namely MPC control, BDS1 and BDS2, were simulated, assuming the same effluent quality and flow rate pattern recorded for the first control experiment described in paragraph 3.1. Simulations were performed for each of the four discharge limits of EU regulation described in paragraph 3. Table 1 reports PAA consumption, dosage and operating conditions for the considered combinations of dosage strategies and discharge limits. In case of the two strictest discharge limits of 10 and 100 CFU/100mL, a significant saving of disinfectant was estimated in case of MPC control, with respect to BDS1 and BDS2. Savings ranged between 30% and 85%, which includes the roughly 50% saving recorded by Manoli et al. (2019) for PAA disinfection to guarantee a concentration of *E. coli* less than 100 CFU/100mL. Importantly, MPC control avoids very high concentrations of PAA, significantly reducing average PAA dosage and applying high dosages only in case of low contact time, high PAA decay and/or high initial *E. coli* concentration. This results in a reduced duration of events with potential ecotoxicological impact due to excessive concentrations of disinfectant at discharge. MPC control was still beneficial with respect to BDS2 in case of a 1000 CFU/mL discharge limit, with a 70% save of PAA. Differently, in the remaining cases, MPC control implied higher consumption of PAA, but with a relevant increase only with respect to BDS1 when the discharge limit is 10,000 CFU/100mL.

Overall, MPC control resulted especially beneficial at discharge limits of 10 to 1000 CFU/mL, particularly with respect to BDS2, which is the conservative traditional control strategy most likely to be used by water utilities. Differently, BDS1 approach, which can be seen as a risk neutral attitude, implied lower PAA consumption than MPC control for higher discharge limits of 1000 and 10,000 CFU/100mL. However, still BDS1 assumes average operating conditions when designing dosage, admitting a high probability of the occurrence of events which are less favorable than nominal conditions. This assumption does not necessarily fit water reuse practices, where tolerable probability of non-compliance is often low.

Table 1 – Consumption of PAA and dosage concentration of PAA (in parenthesis) during control simulation at discharge limits defined by EU regulation.

Dosage strategy	**HRT**	**N_0**	**OD**	**k**	**PAA consumption (mg) (PAA dosage (mg L^{-1}))**			
					≤10	≤100	≤1000	≤10,000
Flow-paced (BDS1)	24	134853	0.05	-0.019	293 (15)	137 (7)	21.5 (1.1)	9.8 (0.5)
Flow-paced (BDS2)	18	523000	0.05	-0.019	489 (25)	293 (15)	88.0 (4.5)	21.5 (1.1)
MPC control	18-46				206 (9.8±4.6)	43.85 (2.2±0.8)	27 (1.3±1.1)	23 (1.1±1.5)

6.4 Conclusions

In this work, feedforward MPC control of wastewater disinfection by PAA was conceived and deployed at pilot scale. This is the first time that real-time control is used in chemical disinfection while considering it as a dynamic system with time-varying hydraulics and kinetic parameters. Importantly, the effectiveness of MPC control in meeting target *E. coli* discharge limit was supported by results of pilot scale experiments, with only a few violations due to hardware failure. PAA concentration variations at the outlet of the disinfection reactor were well predicted, at varying dosage, flow rate and PAA decay. The integration of the PAA disinfection model in the control algorithm is then important to provide good predictions of residual PAA, to understand when potential ecotoxic events occur and, eventually, when quenching of residual disinfectant should be applied.

Simulations of MPC control considering discharge limits of the EU regulation for direct water reuse in agriculture were performed and they estimated significant potential disinfectant saving of 30 to 85% with respect to traditional risk neutral and conservative *flow-paced* dosage strategies, which assume constant PAA concentration. Overall, higher savings were estimated for lower discharge limits, i.e. 10 and 100 CFU/100mL. Therefore, MPC control of disinfection could be really impacting in water reuse applications where high process efficiencies are required, both

resulting in higher reliability, since real-time control compensate time-varying operating conditions, and in much lower disinfectant consumption.

This work paves the way to important future research on robustness in disinfection control. Due to intrinsic variability of microorganism concentration in water and measurement results, *E. coli* concentration at the reactor outlet during pilot control experiment with a setpoint of 10 CFU/100mL was very close to the target, but still concentration slightly passed the limit in some cases, as this MPC control formulation is still deterministic. This result highlights the need for research of the proper robust MPC control formulation to guarantee compliance with a tolerable probability of limit violations.

References

Angelakis, A. N., Asano, T., Bahri, A., Jimenez, B. E., & Tchobanoglous, G. (2018). Water reuse: From ancient to modern times and the future. Frontiers in Environmental Science, 6(MAY). https://doi.org/10.3389/fenvs.2018.00026

Balachandran, S., Charamba, L. V. C., Manoli, K., Karaolia, P., Caucci, S., & Fatta-Kassinos, D. (2021). Simultaneous inactivation of multidrug-resistant Escherichia coli and enterococci by peracetic acid in urban wastewater: Exposure-based kinetics and comparison with chlorine. Water Research, 202(July), 117403. https://doi.org/10.1016/j.watres.2021.117403

Baum, R., Luh, J., & Bartram, J. (2013). Sanitation: A global estimate of sewerage connections without treatment and the resulting impact on MDG progress. Environmental Science and Technology, 47(4), 1994–2000. https://doi.org/10.1021/es304284f

Block, P., Morgan, S., Bell, K., & Stewart, S. (2015). Control strategies for PAA wastewater disinfection at WWtPS with variable effluent quality. 88th Annual Water Environment Federation Technical Exhibition and Conference, WEFTEC 2015, 11, 508–527. https://doi.org/10.2175/193864715819555922

Camacho, E. F., & Bordons, C. (1999). Camacho, Bordons - Model Predictive Control.pdf.

Dieu, B., Garrett, M. T., Ahmad, Z., & Young, S. (1995). Applications of automatic control systems for chlorination and dechlorination processes in wastewater treatment plants. ISA Transactions, 34(1), 21–28. https://doi.org/10.1016/0019-0578(94)00041-J

Domínguez Henao, L., Cascio, M., Turolla, A., & Antonelli, M. (2018). Effect of suspended solids on peracetic acid decay and bacterial inactivation kinetics: Experimental assessment and definition of predictive models. Science of the Total Environment, 643, 936–945.

https://doi.org/10.1016/j.scitotenv.2018.06.219

Domínguez Henao, L., Delli Compagni, R., Turolla, A., & Antonelli, M. (2018). Influence of inorganic and organic compounds on the decay of peracetic acid in wastewater disinfection. Chemical Engineering Journal, 337(October 2017), 133–142. https://doi.org/10.1016/j.cej.2017.12.074

Ducoste, J., Carlson, K., & Bellamy, W. (2001). The integrated disinfection design framework approach to reactor hydraulics characterization. Journal of Water Supply: Research and Technology - AQUA, 50(4), 245–261. https://doi.org/10.2166/aqua.2001.0021

Eaton, J. W., & Rawlings, J. B. (1992). Model-Predictive Control of Chemical Processes. Chemical Engineering Science, 47(4), 705–720. https://doi.org/10.1016/0009-2509(92)80263-C

Foschi, J., Turolla, A., & Antonelli, M. (2021). Soft sensor predictor of E. coli concentration based on conventional monitoring parameters for wastewater disinfection control. Water Research, 191, 116806. https://doi.org/10.1016/j.watres.2021.116806

Han, H. G., Qian, H. H., & Qiao, J. F. (2014). Nonlinear multiobjective model-predictive control scheme for wastewater treatment process. Journal of Process Control, 24(3), 47–59. https://doi.org/10.1016/j.jprocont.2013.12.010

John C. Crittenden, R. Rhodes Trussell, David W. Hand, K. J. H. and G. T. (2017). MHW´s Water Treatment Principles and Design.

Khawaga, R. I., Abdel Jabbar, N., Al-Asheh, S., & Abouleish, M. (2019). Model identification and control of chlorine residual for disinfection of wastewater. Journal of Water Process Engineering, 32(April), 100936. https://doi.org/10.1016/j.jwpe.2019.100936

Kobylinski, E. A., Hunter, G. L., & Shaw, A. R. (2014). On Line Control Strategies for Disinfection Systems: Success and Failure. Proceedings of the Water Environment Federation, 2006(5), 6371–6394. https://doi.org/10.2175/193864706783761716

Lavrnić, S., Zapater-Pereyra, M., & Mancini, M. L. (2017). Water Scarcity and Wastewater Reuse Standards in Southern Europe: Focus on Agriculture. Water, Air, and Soil Pollution, 228(7). https://doi.org/10.1007/s11270-017-3425-2

Lyu, S., Chen, W., Zhang, W., Fan, Y., & Jiao, W. (2016). Wastewater reclamation and reuse in China: Opportunities and challenges. Journal of Environmental Sciences (China), 39, 86–96. https://doi.org/10.1016/j.jes.2015.11.012

Manoli, K, Sarathy, S., Maffettone, R., & Santoro, D. (2019). Advanced control strategies for chemical-based disinfection processes in municipal wastewater : modelling and validation studies.

Manoli, Kyriakos, Sarathy, S., Maffettone, R., & Santoro, D. (2019). Detailed modeling and advanced control for chemical disinfection of secondary effluent wastewater by peracetic acid. Water Research, 153, 251–262. https://doi.org/10.1016/j.watres.2019.01.022

Naidoo, S., & Olaniran, A. O. (2013). Treated wastewater effluent as a source of microbial pollution of surface water resources. International Journal of Environmental Research and Public Health, 11(1), 249–270. https://doi.org/10.3390/ijerph110100249

Onyango, L. A., Quinn, C., Tng, K. H., Wood, J. G., & Leslie, G. (2015). A Study of Failure Events in Drinking Water Systems as a Basis for Comparison and Evaluation of the Efficacy of Potable Reuse Schemes. Environmental Health Insights, 9(s3), 11–18. https://doi.org/10.4137/EHI.S31749

Radcliffe, J. C. (2022). Current status of recycled water for agricultural irrigation in Australia, potential opportunities and areas of emerging concern. Science of the Total Environment, 807, 151676. https://doi.org/10.1016/j.scitotenv.2021.151676

Ritter, W. (2021). State regulations and guidelines for wastewater reuse for irrigation in the U.S. Water (Switzerland), 13(20). https://doi.org/10.3390/w13202818

Rock, C. M., Brassill, N., Dery, J. L., Carr, D., McLain, J. E., Bright, K. R., & Gerba, C. P. (2019). Review of water quality criteria for water reuse and risk-based implications for irrigated produce under the FDA Food Safety Modernization Act, produce safety rule. Environmental Research, 172(December 2018), 616–629. https://doi.org/10.1016/j.envres.2018.12.050

Rodriguez, D., Van Buynder, P., Lugg, R., Blair, P., Devine, B., Cook, A., & Weinstein, P. (2009). Indirect potable reuse: A sustainable water supply alternative. International Journal of Environmental Research and Public Health, 6(3), 1174–1209. https://doi.org/10.3390/ijerph6031174

Santín, I., Pedret, C., Vilanova, R., & Meneses, M. (2015). Removing violations of the effluent pollution in a wastewater treatment process. Chemical Engineering Journal, 279, 207–219. https://doi.org/10.1016/j.cej.2015.05.008

The European Parliament and the Council of the European Union. (2020). Regulation (EU) 2020/741 of 25 May 2020 on minimum requirements for water reuse. Official Journal of the

European Union, 2019(April), L 177/32-L 177/55.

USEPA. (2006). National Primary Drinking Water Regulations: Long term 2 Enhanced Surface Water Treatment Rule (40 CFR Parts 9, 141, and 142). Federal Register, 71(3), 654–786.

Zeng, J., & Liu, J. (2015). Economic model predictive control of wastewater treatment processes. Industrial and Engineering Chemistry Research, 54(21), 5710–5721. https://doi.org/10.1021/ie504995n

Chapter 7: Quantitative Microbial Risk Assessment for risk-based management of ultraviolet disinfection of wastewater

Abstract

Wastewater reuse for irrigation of food crops has a great potential in reducing the water footprint of agriculture, but it requires high efficiencies in removal of enteric pathogens, which are a threat to health of crop consumers and farm workers. In this context, disinfection treatment assumes primary importance and needs to be controlled to reduce health risk coming from wastewater reuse under a tolerable limit. In this work, the indirect agricultural wastewater reuse case study of a wastewater treatment plant (WWTP) in Milan (Italy) was taken as an example of how Quantitative Microbial Risk Assessment (QMRA) can be used to support risk-based management of disinfection when stringent limits for reuse have to be matched. The probabilistic QMRA framework was applied to model fate of reference pathogens (*norovirus*, *salmonella Typhimurium*) from raw wastewater to human receptors, considering WWTP removal (including a UV disinfection stage) , natural attenuation in agricultural canals and final exposure by accidental ingestion of water and soil particles (farm workers) and crop consumption (consumers). Results of risk assessment highlighted the fundamental importance of UV disinfection in controlling health risk, but revealed that additional inactivation phenomena occurring during transport in canals and in agricultural fields provide significant additional barriers to pathogens before human exposure. The study highlighted how QMRA could be used to optimally manage UV disinfection as a component of an integrated wastewater reuse system, completing already existing inactivation barriers while avoiding waste of energy due to excessive UV dose. In fact, operating UV dose could be reduced with respect to business-as-usual practice in the WWTP, with 33% to 66% energy, while maintaining tolerable health risk (> 10^{-6} DALY per person per year, DALY: Disability Adjusted Life Years). Sensitivity analysis of the exposure assessment model used within QMRA framework revealed the importance of data, assumptions and parametrizations used both in modeling the concentration of pathogens at the inlet and throughout the WWTP, and in dose-response modeling, providing indications on most critical aspects of current risk assessment, where future research should focus.

The research work presented in this chapter was carried out during a research stay period of 6 months at the Michigan State University (Michigan, USA). The research work was carried out with the valuable supervision of Prof. Jade Mitchell (MSU) and advisory of Prof. Joan B. Rose (MSU).

This chapter will be submitted soon for publication.

7.1 Introduction

Reuse of treated wastewater in agriculture is spreading worldwide in last decades (Minhas et al., 2022), pushed by national and international authorities (Rock et al., 2019) as a powerful strategy to reduce agriculture water footprint and compensate water shortage in vulnerable areas (Tabatabaei et al., 2020, Ungureanu et al., 2020, Fito & Van Hulle, 2021). Natural drawback of wastewater reuse is the presence of enteric pathogens, even after treatment and at low concentration, which are a source of potential health risk for crop consumers, farmers and anyone who is directly or indirectly exposed to reused wastewater (Mara et al., 2007, Purnell et al., 2020, Kesari et al., 2021, Ofori et al., 2021). Regulations and guidelines for WWTP discharge limits for wastewater reuse are ever more moving to a risk-based approach, with an integrated view of the water reuse system, promoting dynamic discharge limits and proactive assessment and management of health risk, to comply with tolerable health burden thresholds (The European Parliament and the Council of the European Union, 2020, Natural Resources Management Ministerial Council, 2006, World Health Organization, 2006). This leads to important advantages. In fact, after WWTP discharge point, pathogens can decay or regrow while crossing several natural and artificial environments before human exposure, which change according to the water reuse system under study, especially in case of indirect reuse, where treated effluent is diluted and/or transported by an artificial or environmental water body before reuse and human exposure. We need then site-specific risk assessment, integrating description of all phenomena undergone by pathogens in their fate before human exposure, to estimate actual risk and set adequate WWTP efficiency for pathogen removal. This approach is even more important in *de facto* wastewater reuse cases, where unplanned reuse occurs as result of the incidental presence of treated wastewater in water supply sources. In such cases, sources of contamination and fate and transport of wastewater can be more uncertain and strictly influenced by the involved natural and anthropic system and, thus, site-specific risk assessment of the system receiving reclaimed wastewater could reveal the need for *ad-hoc* discharge limits. Moreover, an integrated risk assessment is the basis to set WWTP treatment

efficiencies according to target reduction of pathogens concentration, instead of fecal indicators, explicitly accounting for the actual primary sources of health risk. While being easier and cheaper to monitor, fecal microbial indicators, like fecal coliforms or coliphages, were several times pointed out as ineffective proxies of enteric pathogens (Harwood et al., 2005).

Quantitative Microbial Risk Assessment (QMRA) (Haas et al., 2014) is one of the best-established approaches for assessment of health risk coming from exposure to pathogens and it is adopted and recommended by main wastewater reuse guidelines. QMRA was extensively applied to agricultural wastewater reuse studies, highlighting pros and cons of this practice and providing useful models to explore different exposure scenarios (*inter alia*: Masciopinto et al., 2020, Farhadkhani et al., 2020). However, still we lack of a comprehensive QMRA study of the whole water reuse system, comprising WWTP, fate and transport to the point of exposure and final exposure route, with most of the studies setting the upper system boundary at the outlet of WWTPs (e.g: Ayuso-Gabella et al., 2011, Farhadkhani et al., 2020, Moazeni et al., 2017). Particularly, impact of wastewater disinfection on health risk was not extensively studied with a quantitative approach, despite this is right the purpose of this treatment and what make its cost worth. Moreover, it is rare to find in the literature QMRA of water reuse systems which reports all details and assumption underpinning risk assessment, which would be fundamental to track back the impact of such modeling assumptions on risk and burden of disease estimates.

In this work, QMRA was applied to an indirect agricultural reuse case study in Milan (Italy), where treated effluent from a large scale municipal WWTP is discharged in an agricultural canal to meet irrigation water demand during summer season. QMRA was used to map burden of disease coming from exposure to residual enteric pathogens via accidental ingestion of water during field work and crop consumption. Potential health hazard was mapped over the irrigated agricultural fields and at different doses delivered by the UV disinfection treatment present in the WWTP, to highlight the potential of a risk-based and integrated approach to support optimal management of disinfection in wastewater treatment.

In this work, as according to the QMRA framework (Haas et al., 2014), reference pathogens, being in this case *salmonella* and *norovirus*, were chosen as proxies to model source, fate and transport, exposure and risk of disease for pathogenic enteric bacteria and viruses. As further detailed in paragraph 7.2, literature data on *norovirus* or viral indicators like coliphages were used in this work to model its presence throughout the system understudy, due to lack of site-specific data. Differently, *E. coli* data collected in the system under study were available and they were used as proxies of the presence of pathogen enteric bacteria, expecially during the disinfection process. In fact, *E. coli* is one of the most widespread indicator microorganism for fecal pathogen bacteria,

but it was not considered in this study as a good indicator of viruses too, considering substantial differences in terms of presence in raw wastewater, particle size and resistance to inactivation, which all affect their fate in the WWTP treatment train.

Sensitivity analysis of the QMRA exposure assessment model was carried out to prioritize factors which mostly drive risk of disease and understand how assumptions impacted the final risk (Zhiteneva et al., 2021), with the aim of giving indications to future research and practice on where to focus experimental and modeling efforts.

7.2 Materials and methods

7.2.1 Description of case study and data collection

In the wastewater reuse system under study, the WWTP of Milano S. Rocco (1.050.000 PE), located in south-west Milan (Italy), treats wastewater for indirect reuse in agriculture in a large area of the "Rural Park South Milan" (Parco Agricolo Sud Milano) and the province of the city of Pavia.

S. Rocco WWTP is based on activated sludge process for nitrification and denitrification, followed by sand filtration and UV disinfection. The UV disinfection stage is made of two parallel lines. This study focused on one line, dedicated to treatment of the fraction of effluent which is reused in agriculture. The disinfection line under study is made of three identical parallel open channels, each one equipped with three banks containing 144 monochromatic (254 nm) low pressure/high intensity UV lamps each (378 lamps per line, 1134 in total), for a total of about 136 kW of installed power per channel. In each channel, the UV fluence rate can be adjusted by changing the number of active banks: thus, each channel can work at three discrete levels of nominal UV dose. The system is designed to deliver a dose of 59 mJ cm^{-2} with all UV banks active and a flow rate of 4 m^3 s^{-1}. Flow rate passing through the wastewater reuse disinfection line is controlled according to the demand of irrigation water.

During the irrigation season, which can at most last from June to September, WWTP delivers treated effluent to the two canals Carlesca and Pizzabrasa (from now on referred as canal C and P, respectively, for simplicity). Canal C and P originate from Naviglio Grande canal and the Lambro Meridionale river, respectively, and thus treated effluent is diluted at the point of discharge. The two canals transport water up to 40 km far from the point of discharge. Water is used for flood irrigation of mainly rice and corn fields, which are 47% and 24% of the irrigated area, respectively. Moreover, many traditional, organic and didactic farms with heterogeneous production are present in the area. The case study area is detailed in Figure 1.

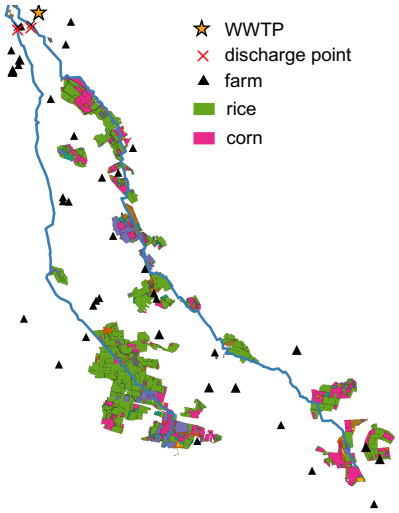

Figure 1 – Map of the case study

7.2.2 QMRA framework

The QMRA framework was adopted to assess risk and burden of disease coming from reuse of treated wastewater in this case study. First, within the Hazard Identification (HI), *norovirus* and *salmonella Typhimurium*, which are both two of the main causes of human gastroenteritis, were chosen as reference for viral and bacterial enteric pathogens, which represents the source of health risk in wastewater reuse practices. Then, the Exposure Assessment, Dose-Response and Risk Characterization phases followed, as detailed in following paragraphs.

7.2.3 Exposure assessment

Concentration of reference pathogens (RP) was modeled from the source, which is raw wastewater at the inlet of the WWTP, to irrigation water. This required modeling of removal throughout WWTP treatment train, dilution, transport and natural attenuation in agricultural canals and final route of exposure from irrigation water to humans. The exposure assessment model was calibrated on published literature data coming from both current case study and other case studies in developed countries, where wastewater characteristics were considered reasonably similar. Since a probabilistic QMRA approach was adopted, variability and uncertainty of model inputs and parameters were described by probability distribution and propagated on risk estimate by means of a Monte Carlo approach. Details on data and sources used for exposure assessment model calibration are reported in Table 1.

Normal, lognormal, exponential and gamma distributions were calibrated and compared via Maximum Likelihood estimation to model *norovirus* and *salmonella* concentration in raw wastewater and their log-removal in activated sludge and sand filtration. Literature data on *norovirus*, *norovirus*-specific surrogates and general virus surrogates from comparable case studies

were used to model *norovirus* concentration. Probability distribution of *norovirus* in raw wastewater was calibrated only on qPCR data on *norovirus* NovGI and NovGII genotype, and the actual concentration of infectious *norovirus* was estimated applying a conversion factor as assumed in previous risk assessment studies (Simhon et al., 2020, Estes et al., 2019). Probability distributions of log-removal of *norovirus* by activated sludge and sand filtration were calibrated on data about various virus species, due to the limited amount of *norovirus*-specific data in the literature. *Norovirus* removal by UV disinfection kinetics was modeled as a log-linear function of UV dose (D), based on several studies on *norovirus*-specific surrogates Murine Norovirus (MNV) and Feline Calicivirus (FCV), whose data come only from culture-based methods. In fact, recent studies highlighted how qPCR data, even if they address the specific virus species under study, are not reliable in detecting inactivation by disinfection (Rönnqvist et al., 2014, Girones et al., 2010, Knight et al., 2016, Stals et al., 2013). Available studies explored *norovirus* UV disinfection at a maximum dose of roughly 40 mJ cm^{-2}, corresponding to a log-removal of approximately 4.5. Thus, the UV disinfection model for *norovirus* extrapolated inactivation at higher doses as the inactivation predicted at 40 mJ cm^{-2}, in order to be conservative and not to overestimate dramatically log-removal. Overall, *norovirus* removal by the WWTP treatment train was modeled as:

$$N_{WWTP} = 10^{(LR_{AS}+LR_{SF}+LR_{UV})} \quad \text{(Eq.1)}$$

where N_{WWTP} is the concentration of *norovirus* removal at the WWTP outlet (CFU/100ml), LR_{AS}, LR_{SF} and LR_{UV} are log-removals (on base 10) in activated sludge, sand filtration and UV disinfection, respectively.

Previously published *E. coli* concentration data on raw wastewater in Milan (Turolla et al., 2017, 2018) and on the current case study (Foschi et al., 2021a) were used to calibrate *salmonella* concentration and log-removal in WWTP. In details, probability distribution of *salmonella* in raw wastewater was modeled as the product between the concentration of *E. coli* and the ratio between *salmonella* and *E. coli* concentration, modeled as a Uniform distribution, based on data coming from Kacprzak et al., 2015. Then, probability distribution of *E. coli* concentration at the outlet of sand filtration was calibrated and used to estimate the log-removal of *salmonella* by activated sludge and sand filtration, assuming that the inactivation of *E. coli* and *salmonella* are equal. Finally, UV disinfection kinetics was modeled as a double exponential function of UV dose, based on data from Foschi et al. (2021). *Salmonella* removal by WWTP treatment train was then modeled as:

$$N_{WWTP} = 10^{(LR_{AS+SF}+LR_{UV})} \quad \text{(Eq. 2)}$$

where LR_{AS+SF} is the cumulated removal by activated sludge and sand filtration.

Table 1 - Summary of details and sources of literature data used to model pathogen concentration and removal in WWTP. (NoVGI and NoVGII: norovirus genotype I and II, EV: enterovirus, AdV: adenovirus, RV: rotavirus, SaV: sapovirus, JC virus, AsV astrovirus, PMMoV: pepper mild mottle virus, polyoma virus, CV: calicivirus, HAV: hepatitis A virus, MNV: murine norovirus, MS2: MS2 coliphage)

WWTP phase	Source	Dataset size	Location	Microorganism	Measurement method
Raw wastewater	Cioffi et al., 2020	1	Italy	NoVGI	qPCR
	Kitajima et al., 2014	24	USA	NoVGI	qPCR
	La Rosa et al., 2010	1	Italy	NoVGI	qPCR
	Cioffi et al., 2020	1	Italy	NoVGII	qPCR
	Kitajima et al., 2014	24	USA	NoVGII	qPCR
	La Rosa et al., 2010	1	Italy	NoVGII	qPCR
	McCall et al., 2020	43	USA	NoVGII	qPCR
	Masclaux et al., 2013	12	Switzerland	NoVGII	qPCR
	Turolla et al., 2017	5	Italy	*E. coli*	culture
	Turolla et al., 2017	5	Italy	*E. coli*	culture
	Turolla et al., 2017	7	Italy	*E. coli*	culture
Removal in activated sludge	Naughton and Rousselot, 2017	18	Various	EV, NoV, AdV, RV, SaV, JCV virus, AsV, poliovirus, BK virus, Torque teno virus, coliphages	qPCR and culture
	Kacprzack et al., 2014	4	Southern Poland	*E. coli, Salmonella Typhimurium**	culture
	Rajala et al., 2003	1	Finland	F-coliphage	culture
	Hijnen et al., 2010	2	Netherlands	F-coliphage	culture
	Shirasaki et al., 2012	1	Japan	F-coliphage	culture
	Asami et al., 2016	4	Thailand	PMMoV, JCV PyV	qPCR
Removal in sand filtration	Hokajärvi et al., 2018	2	Finland	F-coliphage	culture
	Shirasaki et al., 2018	7	Japan	AdV, CV, HAV, MNV, MS2, PhiChi174, PMMoV	qPCR, culture
	Kato et al., 2018	8	Japan	PMMoV, AiV, MNV, NoVGII, MS2, Qbeta	qPCR, culture
	WHO	1		F-coliphage	culture
Removal in activated sludge and sand filtration	Foschi et al., 2021		Italy (this case study)	*E. coli*	culture

Dilution of treated effluent after discharge was accounted by mass balance, assuming perfect mixing. Agricultural canals were assumed to behave as ideal plug-flows with space-invariant speed and uniform section. Natural attenuation of *norovirus* and *salmonella* along the canals was then estimated assuming a first order decay kinetics. Probability distributions of first order decay rates

were for *norovirus* and *salmonella* calibrated on data coming from comprehensive recent reviews on pathogen persistence in surface water (Boehm et al., 2018, 2019). As for previously described parameters, gaussian, lognormal, exponential and gamma distributions were compared. Concentration of pathogens along canals C and P were then calculated as:

$$N_{CAN} = N_{WWTP} e^{-k\frac{x}{v}} \quad \text{(Eq. 3)}$$

where N_{CAN} is the concentration of pathogen at distance x (km) from the discharge point, k is the natural inactivation first order decay rate (d^{-1}) and v is the flow velocity in the canal. Concentration of pathogens in irrigation water in a given agricultural field was assumed to be equal to N_{CAN} in the closest point of the canal.

Two exposure scenarios were considered. The first scenario (scenario 1) considers farmers and field workers involuntary ingesting water and soil particles while working in fields irrigated by flood irrigation. A previously developed model (Symonds et al., 2014, Moazeni et al., 2017) was used to simulate the volume of water and soil accidentally ingested by person per day. The model assumes that both ingestion of water and soil can be described equivalently as uniform probability distribution comprised between 0.1 and 1 ml d^{-1}. In case of soil, an additional 2 log inactivation was considered. The exposure dose d (CFU day^{-1} $pers.^{-1}$ or PFU day^{-1} $pers.^{-1}$) in scenario 1 was then estimated as:

$$V_w \sim U(0.1,1) \quad \text{(Eq. 4)}$$

$$d_s = C_{CAN} V_w 10^{-k_{soil}} \quad \text{(Eq. 5)}$$

$$d_w = N_{CAN} V_w \quad \text{(Eq. 6)}$$

$$d = d_s + d_w \quad \text{(Eq. 7)}$$

where V_w is the amount of water (or water contained in soil) ingested (ml $pers.^{-1}$ d^{-1}), k_{soil} is pathogen log-removal in soil, d_s and d_W are the pathogen doses ingested from water and soil, respectively.

The second exposure scenario (scenario 2) considers the consumption of crops irrigated with treated wastewater. Since the case study contains many organic farms, consumption of lettuce was considered as worst-case scenario. Model developed by Mok and Hamilton (2014) was used to estimate exposure dose, as described by following equations:

$$d = N_{CAN} V_R c e^{-k_I T} 10^W \quad \text{(Eq. 8)}$$

$$V_R \sim Lognormal3(-4.57, 0.50, 0.006) \quad \text{(Eq. 9)}$$

$$k_I \sim Normal(1.07, 0.07) \text{ (truncated at 0)} \quad \text{(Eq. 10)}$$

$$W \sim PERT(0.1, 1, 2) \quad \text{(Eq. 11)}$$

where V_R is the volume of water retained by lettuce crop surface (ml g^{-1}), c (g pers.$^{-1}$ d.$^{-1}$) is lettuce consumption in Italy (Leclercq et al., 2009), k_I is infield decay on lettuce surface (d^{-1}), T (d) is the time between last irrigation and harvest, and W is log-removal during lettuce washing.

The two exposure scenarios were evaluated at 4 different UV disinfection dose levels, corresponding to 0, 1, 2 and 3 active banks of UV lamps per channel (see paragraph 2.1). In all the combinations of exposure scenarios and number of active banks, the scenario of irrigation water demand from 2015 was considered, since it was the highest in the last decade for the case under study. Average flow rate of treated wastewater to be delivered to agriculture was 2 m^3 s^{-1}. Variability of flow rate was accounted in Monte Carlo simulation by randomly sampling daily flow rate from data records from 2015.

7.2.4 Dose-response model

Dose-response model developed by Teunis et al. (2008) was used to estimate risk of disease per person per day from *norovirus*. Model assuming no aggregation of viral particles was adopted, being the most conservative option, as highlighted in Van Abel et al. (2017). Risk of disease from *norovirus* was then calculated as:

$$P(inf) = 1 - {}_1F_1(\alpha, \beta, -d) \quad \text{(Eq. 12)}$$

$$P(dis|inf) = 1 - (1 + \eta d)^{-r} \quad \text{(Eq. 13)}$$

$$P(dis) = P(inf)P(dis|inf) \quad \text{(Eq. 14)}$$

where ${}_1F_1(\alpha, \beta, -d)$ is the hypergeometric function with parameters α and β, $P(inf)$ is the probability of infection, $P(dis|inf)$ is the probability of disease given infection with parameters η and r and $P(dis)$ is the probability of disease. All probabilities are meant as per person per day.

Probability of disease from *salmonella* was estimated by dose-response model derived from outbreak data on non-typhoidal *salmonella* and assuming a Beta-Poisson model (Skovgaard, 2004). In the case of *salmonella*, probability of disease is directly estimated as:

$$P(dis) = 1 - \left(1 + \frac{d}{\beta}\right)^{\alpha} \quad \text{(Eq. 15)}$$

Uncertainty in dose-response model parameters was estimated and propagated by means of a Monte Carlo simulation by the bootstrap method (Weir et al., 2017).

Annual risk of disease was then computed as:

$$P_Y = 1 - \left(1 - P(dis)\right)^{n} \quad \text{(Eq. 16)}$$

assuming n as equal to 120 days, since the irrigation season can last up to 4 months in the case study.

7.2.5 Risk characterization

Burden of disease was estimated using Disability Adjusted Life Years (DALY) as indicator, computed as:

$$DALY = YLL + YLD \quad \text{(Eq. 17)}$$

$$YLL = mL_{exp}A_d \quad \text{(Eq. 18)}$$

$$YLD = sd \quad \text{(Eq. 19)}$$

where YLL are "years of life lost", YLD are "years of life with disability", L_{exp} is life expectancy at the age of death, m is mortality, A_d is the age distribution of deaths for the given illness, s is severity (or disability weight) and d is the duration of disease. Assumptions and sources for DALY calculations are reported in Tables 2 and 3.

Table 2 – DALY model parameters assumed in this study, with sources and main underpinning assumptions.

Variable	Value		Source	Assumptions
	Salmonella	***Norovirus***		
m	0.0017	0.0017	http://ghdx.healthdata.org Global Burden of Disease Study 2019 (GBD 2019)	– data referred to Italy in 2019 – uniform for all age classes – the 95th percentile was assumed
s	0.247	0.247	http://ghdx.healthdata.org Global Burden of Disease Study 2019 (GBD 2019)	– global data – uniform for all age classes – referred to "acute gastroenteritis"
d	0.0296	0.0153	Kemmeren et al., 2006	– data referred to the Netherlands and UK

Table 3 – Summary of assumptions and sources on age distribution of deaths due to gastroenteritis and life expectancy at the age of death used for DALY calculations.

Age class	***Salmonella***		***Norovirus***	
	Proportion of deaths[1]	**Life expectancy at the age of death[2]**	**Proportion of deaths[1]**	**Life expectancy at the age of death[2]**
0-4 years	0.265	80.22	-	-
0-5 years	-	-	81.979	0.05
5-9 years	0.077	75.259	-	-
10-14 years	0.068	70.283	-	-
15-64 years	0.342	42.765	-	-
>64 years	-	-	20.729	0.95
>65 years	0.248	19.898	-	-

[1] Kemmeren et al., 2006
[2] Italian National Statistical Institute (ISTAT)

7.2.6 Sensitivity analysis

Correlation-based sensitivity analysis of the exposure assessment model, considering risk of disease per person per day as model output, was performed to prioritize model inputs and parameters. A Monte Carlo simulation of the model (n = 10000) was run, randomly sampling model

input and parameters from their probability distributions, to explore model sensitivity in a realistic space. For each input or parameter, Spearman correlation coefficient was computed. Sensitivity analysis was run separately for *norovirus* and *salmonella*, for the two exposure scenarios and for the 4 UV doses considered.

7.3 Results and discussion

7.3.1 Risk assessment

Parameters of WWTP treatment and natural attenuation in canals, which were specifically calibrated for this study from literature data, are summarized in Table 4.

Probabilistic QMRA was performed via Monte Carlo simulation under the two exposure scenarios described in paragraph 2.2.1, considering in each case 4 UV disinfection scenarios, at increasing UV dose. Results of risk assessment in all of the 8 combinations of exposure scenario and UV dose were compared with the limit of tolerable burden of disease adopted by many guidelines for water reuse (WHO, Australian Guidelines, EU regulation), which is equal to a 10^{-6} DALY per person per year. Table 5 reports results on estimated DALY in case human exposure occurs immediately after the point of discharge and dilution in agricultural canals, as worst-case scenario where natural attenuation of pathogens is null. QMRA revealed the fundamental importance of the presence of UV disinfection in the WWTP. In fact, both in case of accidental ingestion of water and crop consumption, average DALY is significantly higher than the tolerable threshold, when no UV disinfection is applied. Low and middle UV doses are sufficient to lower the probability of passing the threshold to negligible values for all scenarios and pathogens, except for *salmonella* in case of scenario 1. The increase of UV dose to the highest level did not significantly reduce DALY level.

DALY was then mapped over the agricultural fields, to assess the combined effect of UV disinfection and natural attenuation along agricultural canals in lowering risk. As an example of interpreting QMRA results with the most conservative approach, Figure 2 reports trend of 95th percentile of DALY with respect to the distance from the discharge point and at increasing UV dose and it is compared with the tolerable threshold of 10^{-6} DALY. As can be seen, in case of *norovirus* almost the whole reduction in risk is obtained by low dose UV disinfection, while the effect of further increasing UV dose and distance from the discharge point is negligible. Differently, increase of UV dose to middle level and natural attenuation have a significant role in case of risk from *salmonella*, especially in case of scenario 1. In fact, even if at a given dose the probability of having a non-tolerable DALY at the point of discharge is remarkable, the inclusion of natural attenuation

model within the QMRA exposure assessment allow us to identify agricultural areas which are sufficiently far from the WWTP to be irrigated safely, as exemplified in Figure 3, where risk from *salmonella* is considered for both scenario 1 and 2 at increasing UV dose.

Reported effects of UV dose on health risk control are significantly influenced by data and models assumed for *norovirus* and *salmonella* UV disinfection kinetics, as reported in Figure 4. Inactivation kinetics of *norovirus* showed higher rate and a higher maximum log removal in the UV dose range under study. Differently, the inactivation model used for *salmonella* shows a sharp tailing effect. This could be due to the differences in data sources: due to lack of data, *norovirus* inactivation was calibrated on *norovirus*-specific surrogates in laboratory experiments on synthetic water matrices. Differently, inactivation model for *salmonella* inactivation was calibrated on *E. coli* data collected in continuous flow full-scale disinfection experiments carried out in the disinfection unit under study. It is then likely that the tailing effect recorded for bacteria inactivation better represents inactivation kinetics in real conditions.

Table 4 – Summary of results of calibration of models for concentration of pathogens in raw wastewater, removal in WWTP and natural attenuation in surface water.

Variables/ processes	Units	Models and parameters
WWTP - *norovirus*		
Concentration in raw wastewater	PFU/100 ml	NovGI: ~Lognormal(10.53, 1.91) NovGII: ~Lognormal(10.82, 2.08)
Log-removal in activated sludge	-	~Normal(1.89, 0.76)
Log-removal in sand filtration	-	~Lognormal(-0.35, 0.64)
Log-removal in UV disinfection	-	$LR_{UV} = -k_N D$ k_N=0.14
WWTP - *salmonella*		
Concentration in raw wastewater	CFU/100ml	E. coli: ~Lognormal(14.64, 0.62) R_{S-EC}~U(0.02, 1.02) $N_0 = EC R_{S-EC}$
Removal in activated sludge and sand filtration[1]	-	N_{out}~ Gamma(85.66,10.24) $LR_{SF} = \frac{N_{out}}{N_{RW}}$
Removal in UV disinfection	-	$N_{out} = \delta N_0 e^{-k_{S,1}D} + (1-\delta) N_0 e^{-k_{S,2}D}$ δ=0.999; $k_{S,1}$=0.17; $k_{S,2}$=0.005 $LR_{UV} = log_{10} \frac{N_{out}}{N_0}$
Transport and natural attenuation in agricultural channels		
Natural inactivation of *norovirus*	d^{-1}	~Gamma(0.92, 8.13)
Natural inactivation of *salmonella*	d^{-1}	~Lognormal(-0.59, 0.74)

[1]meant as cumulated removal of the sequence of the two treatments

Currently, UV disinfection in the considered case study is run at maximum UV dose, in order to reliably comply with the Italian regulation for wastewater reuse in agriculture, which is highly conservative and defines a strict discharge limit on the indicator microorganism *E. coli* equal to 10 CFU/100ml, regardless the characteristics of the downstream water reuse system, the irrigation

system and the type of irrigated crop. Such conservative management of UV disinfection implies a consumption of roughly 870 MWh, which, according to recent years WWTP records, could reach a share of up to 12% of the total energy consumption. Risk assessment results revealed that, even according to a very conservative approach, irrigation water would be safe in almost all considered scenarios, with the exception of the accidental ingestion of *salmonella* by workers in agriculture (see Figures 2 and 3). However, even in such cases, estimated probability of passing the tolerable DALY threshold is very low. Moreover, the portion of agricultural areas identified as not safe is limited to a distance lower than 5.5 km for middle UV dose and 15 km for low UV dose. In such cases, the water reuse system could be managed by avoiding proximity to water during irrigation events. UV disinfection could be then run at low or middle UV dose, with a potential saving of 66% or 33%, respectively.

Table 5 – Average DALY (± std. deviation) at point of discharge estimated by QMRA. Probability of having DALY $>10^{-6}$ is reported in parenthesis.

Average UV dose [mJ cm^{-2}]	**Energy consumption [MWh]**	**Scenario 1**		**Scenario 2**	
		norovirus	***salmonella***	***norovirus***	***salmonella***
0	0	3.6×10^{-3} $\pm8.7\times10^{-3}$ (88.1%)	9.7×10^{-3} $\pm1.6\times10^{-2}$ (99.1%)	1.0×10^{-4} $\pm9.3\times10^{-4}$ (42.7%)	7.5×10^{-5} $\pm5.4\times10^{-4}$ (80.9%)
45	290	1.9×10^{-7} $\pm5.0\times10^{-6}$ (<1%)	4.2×10^{-5} $\pm2.3\times10^{-4}$ (60.9%)	4.8×10^{-10} $\pm8.5\times10^{-9}$ (<1%)	1.3×10^{-7} $\pm5.6\times10^{-7}$ (3.1%)
90	580	1.37×10^{-8} $\pm3.6\times10^{-7}$ (<1%)	7.24×10^{-7} $\pm2.3\times10^{-6}$ (12.6%)	1.36×10^{-9} $\pm3.8\times10^{-8}$ (<1%)	2.1×10^{-8} $\pm8.9\times10^{-8}$ (<1%)
135	870	9.2×10^{-9} $\pm1.8\times10^{-8}$ (<1%)	9.0×10-7 ±5.4×10-6 (12.4%)	2.8×10^{-10} $\pm3.5\times10^{-9}$ (<1%)	2.7×10^{-8} $\pm1.3\times10^{-7}$ (<1%)

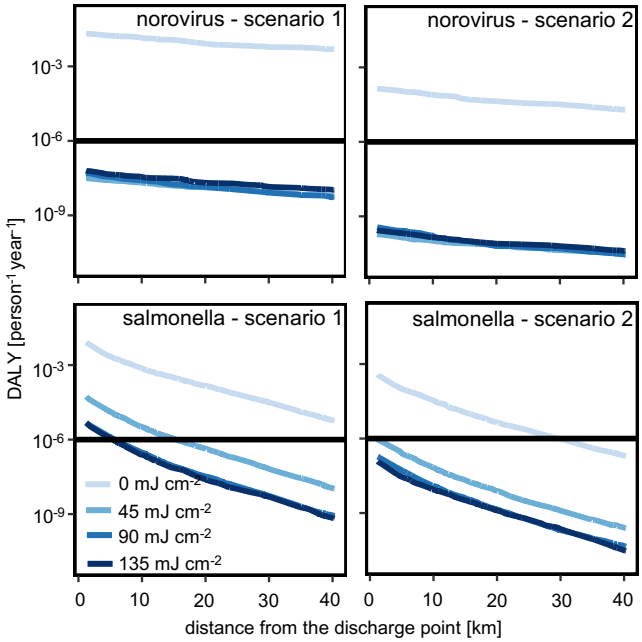

Figure 2 – Trend of 95th percentile of DALY person^{-1} y^{-1} with distance from the discharge point and UV dose.

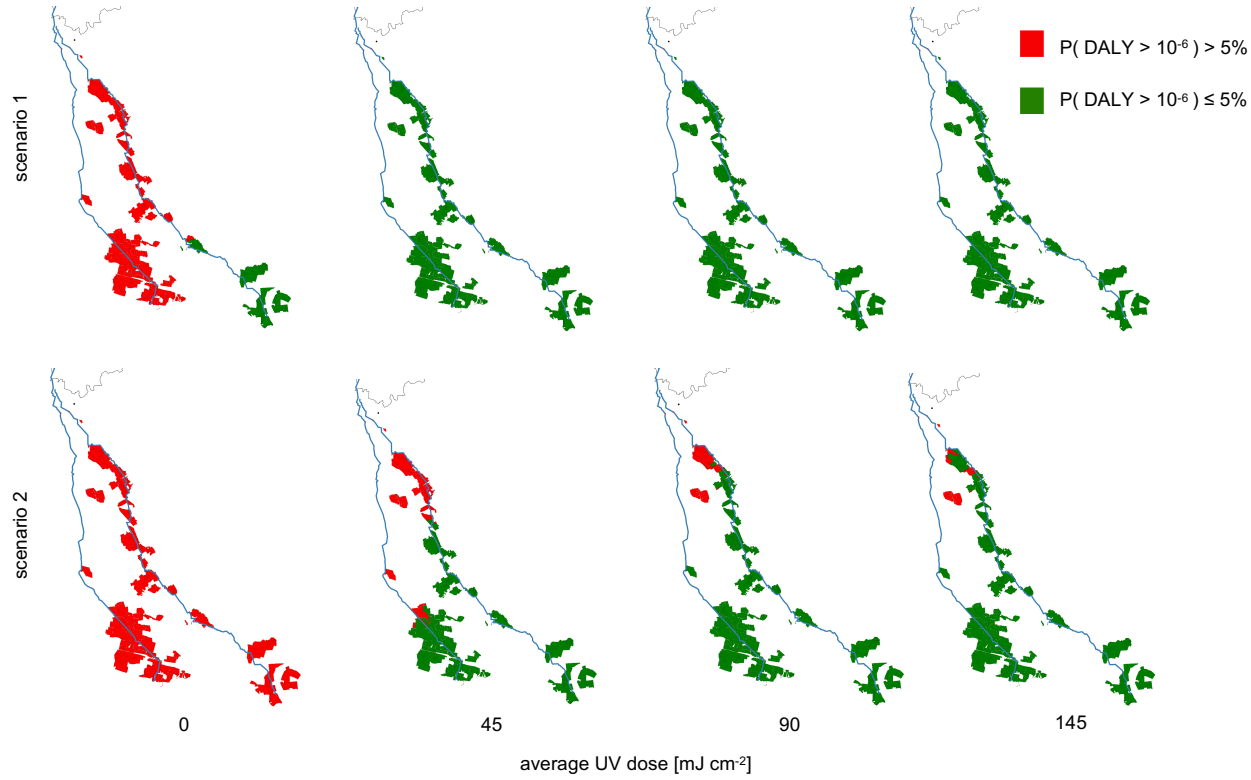

Figure 3 – Mapping of "safe" (green) and "non safe" (red) area for indirect reuse of treated wastewater, identified comparing 95th percentile of DALY person^{-1} y^{-1} with the tolerable threshold equal to 10^{-6}.

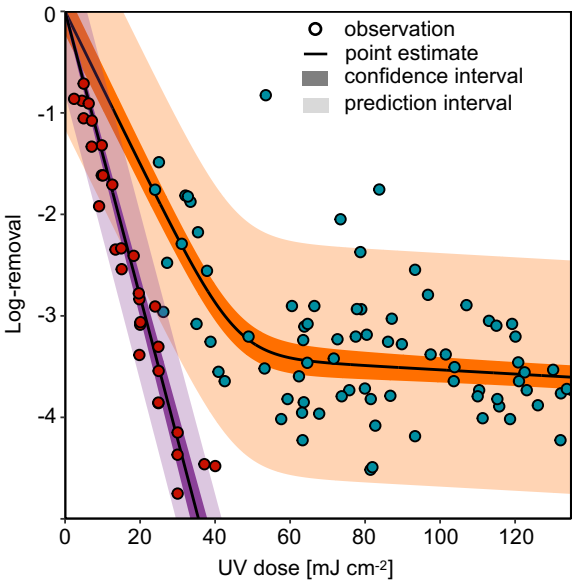

Figure 4 – Results of UV kinetic model fitting to literature data.

7.3.2 Sensitivity analysis

Correlation-based sensitivity analysis of the exposure assessment model was performed to prioritize importance of model inputs and parameters in determining the risk of disease from *norovirus* and *salmonella*. The sensitivity analysis was repeated at each of the 4 levels of UV dose explored in risk assessment, whose results confirmed the importance of disinfection in controlling risk of disease. Results, reported as heatmap in Figure 5, highlight relative importance of the rest of considered model inputs and parameters in driving estimated risk of disease, at varying UV doses and scenarios.

As expected, parameters related to WWTP are among the most important. In details, variability of concentration in raw wastewater and removal in activated sludge and sand filtration show the highest correlation coefficients, especially if disinfection is not active. Differently, uncertainty of UV model parameters is not as relevant. Treated effluent flow rate seems to be quite relevant only in case of *salmonella* and low UV doses. This can be explained by the characteristics of UV disinfection of bacteria. In fact, in this case study as in many others in the literature, bacteria UV disinfection kinetics follows a biphasic behavior, with an initial high-rate phase and a sequent low-rate phase (see paragraph 2.2.1 and Table 4). Low level UV dose in this case study, which is equal to about 45 mJ cm^{-2}, corresponds roughly to the point where the kinetics change from one phase to the other. Small variations of flow rate, resulting in variations of the dose, imply then big and nonlinear variations of inactivation.

As expected by analyzing results from risk assessment, natural inactivation in agricultural canals is more relevant in case of *salmonella*. In fact, *salmonella* first order decay parameter mean value is 0.83 d^{-1}, significantly higher than *norovirus*, whose mean value is 0.13 d^{-1}.

Finally, sensitivity analysis revealed parameters η and r (see paragraph 2.2.1) of the dose-response model of *norovirus* have a big impact in determining the risk of disease. This is due to the high uncertainty in determining those parameters, due to a very limited dataset size on *norovirus* probability of illness given infection, as reported in Teunis et al. (2008).

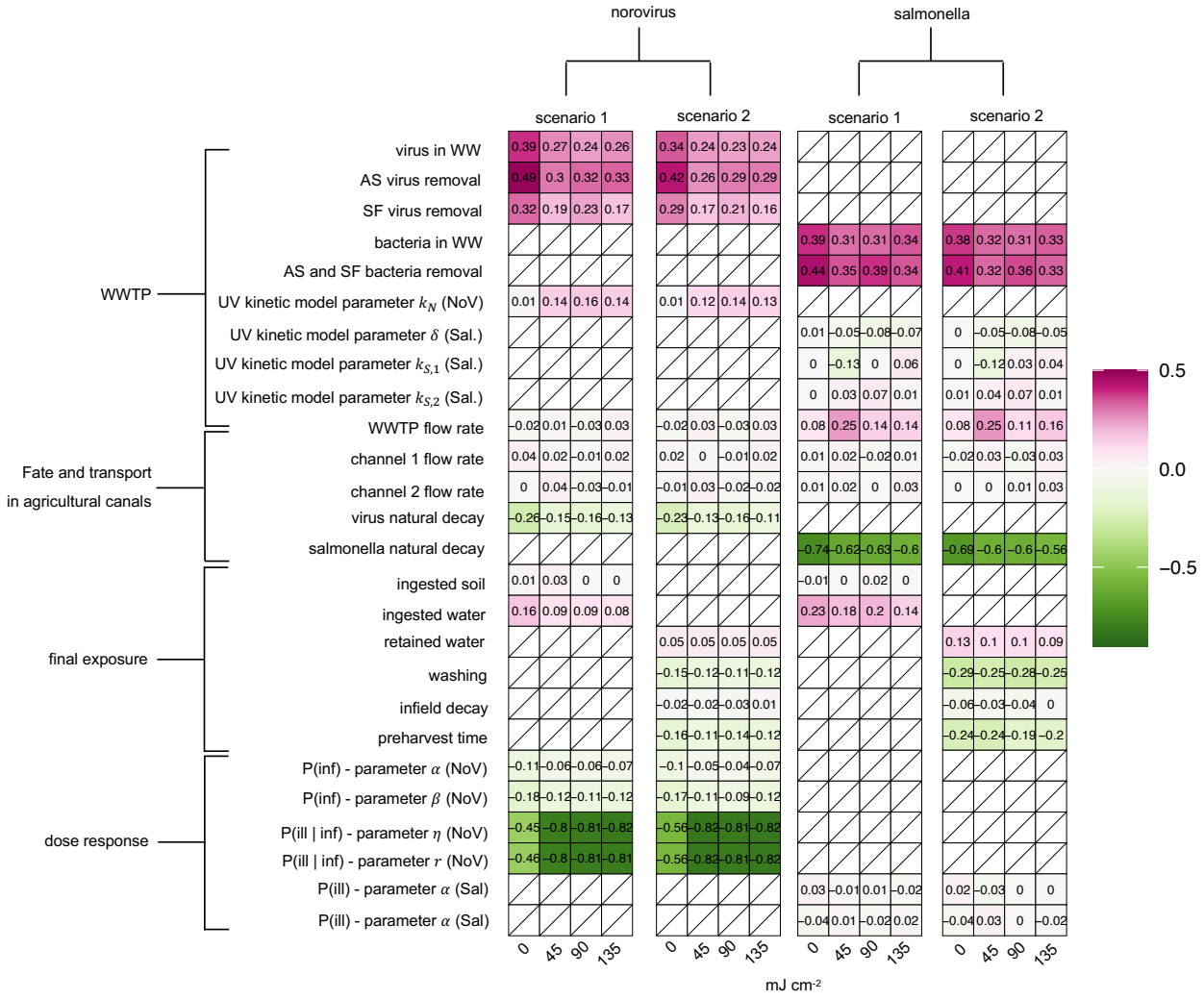

Figure 5 – Heatmap of the correlation-based sensitivity analysis of risk of disease, at varying scenario, pathogen and UV dose. Spearman correlation coefficients are mapped in colors. Tiles are crossed in case model input is not related to the considered pathogen or scenario.

7.4 Conclusions

In this work, QMRA of an indirect wastewater reuse in agriculture was performed, accounting for fate of pathogens from raw wastewater through WWTP removal, natural attenuation in agricultural canals and final exposure of humans due to accidental ingestion of water and soil particles and crop consumptions. The study confirmed that the UV disinfection system of the

WWTP is fundamental to maintain tolerable health risk, but highlighted how significant energy saving could be achieved if UV dose is optimized considering it as an integration of natural additional inactivation barriers due to natural inactivation in canals and final inactivation in fields, conceiving the wastewater reuse system as integrated. UV dose is optimized according to a risk-based approach, which is based on actual characteristics of the reuse system, like the type of irrigation system and crop, and on modeling of fate of pathogen of concern, instead of a static limit at the point of discharge on indicator microorganism.

The sensitivity analysis performed on exposure assessment model revealed that parameters related to many steps of pathogens fate from raw wastewater to humans can drive the estimated risk of disease, pointing at concentrations in raw wastewater, removal in the WWTP train, natural inactivation of *salmonella* and dose-response of *norovirus* as most important factors. This states the importance on the collection of data on pathogen concentrations, instead of indicators, in WWTP and surface waters, to better characterize occurrence and removal according to different strains and operating and environmental conditions. Uncertainty and data scarcity on dose-response models, like for *norovirus*, could be reduced by further research on past outbreaks, like in the case of the dose-response model used for *salmonella* in this work.

References

Ayuso-Gabella, N., Page, D., Masciopinto, C., Aharoni, A., Salgot, M., & Wintgens, T. (2011). Quantifying the effect of Managed Aquifer Recharge on the microbiological human health risks of irrigating crops with recycled water. Agricultural Water Management, 99(1), 93–102. https://doi.org/10.1016/j.agwat.2011.07.014

Boehm, A. B., Graham, K. E., & Jennings, W. C. (2018). Can We Swim Yet? Systematic Review, Meta-Analysis, and Risk Assessment of Aging Sewage in Surface Waters. Environmental Science and Technology, 52(17), 9634–9645. https://doi.org/10.1021/acs.est.8b01948

Boehm, A. B., Silverman, A. I., Schriewer, A., & Goodwin, K. (2019). Systematic review and meta-analysis of decay rates of waterborne mammalian viruses and coliphages in surface waters. Water Research, 164, 114898. https://doi.org/10.1016/j.watres.2019.114898

Estes, M. K., Ettayebi, K., Tenge, V. R., Murakami, K., Karandikar, U., Lin, S. C., Ayyar, B. V., Cortes-Penfield, N. W., Haga, K., Neill, F. H., Opekun, A. R., Broughman, J. R., Zeng, X. L., Blutt, S. E., Crawford, S. E., Ramani, S., Graham, D. Y., & Atmar, R. L. (2019). Human Norovirus Cultivation in Nontransformed Stem Cell-Derived Human Intestinal Enteroid

Cultures: Success and Challenges. Viruses, 11(7), 9–11. https://doi.org/10.3390/v11070638

Farhadkhani, M., Nikaeen, M., Hadi, M., Gholipour, S., & Yadegarfar, G. (2020). Campylobacter risk for the consumers of wastewater-irrigated vegetables based on field experiments. Chemosphere, 251, 126408. https://doi.org/10.1016/j.chemosphere.2020.126408

Fito, J., & Van Hulle, S. W. H. (2021). Wastewater reclamation and reuse potentials in agriculture: towards environmental sustainability. Environment, Development and Sustainability, 23(3), 2949–2972. https://doi.org/10.1007/s10668-020-00732-y

Foschi, J., Turolla, A., & Antonelli, M. (2021). Artificial neural network modeling of full-scale UV disinfection for process control aimed at wastewater reuse. Journal of Environmental Management, 300(May), 113790. https://doi.org/10.1016/j.jenvman.2021.113790

Girones, R., Ferrús, M. A., Alonso, J. L., Rodriguez-Manzano, J., Calgua, B., de Abreu Corrêa, A., Hundesa, A., Carratala, A., & Bofill-Mas, S. (2010). Molecular detection of pathogens in water - The pros and cons of molecular techniques. Water Research, 44(15), 4325–4339. https://doi.org/10.1016/j.watres.2010.06.030

Harwood, V. J., Levine, A. D., Scott, T. M., Chivukula, V., Lukasik, J., Farrah, S. R., & Rose, J. B. (2005). Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. Applied and Environmental Microbiology, 71(6), 3163–3170. https://doi.org/10.1128/AEM.71.6.3163-3170.2005

Kacprzak, M., Fijałkowski, K., Grobelak, A., Rosikoń, K., & Rorat, A. (2015). Escherichia coli and Salmonella spp. Early diagnosis and seasonal monitoring in the sewage treatment process by EMA-qPCR method. Polish Journal of Microbiology, 64(2), 143–148. https://doi.org/10.33073/pjm-2015-021

Kesari, K. K., Soni, R., Jamal, Q. M. S., Tripathi, P., Lal, J. A., Jha, N. K., Siddiqui, M. H., Kumar, P., Tripathi, V., & Ruokolainen, J. (2021). Wastewater Treatment and Reuse: a Review of its Applications and Health Implications. Water, Air, and Soil Pollution, 232(5). https://doi.org/10.1007/s11270-021-05154-8

Knight, A., Haines, J., Stals, A., Li, D., Uyttendaele, M., Knight, A., & Jaykus, L. A. (2016). A

systematic review of human norovirus survival reveals a greater persistence of human norovirus RT-qPCR signals compared to those of cultivable surrogate viruses. International Journal of Food Microbiology, 216, 40–49. https://doi.org/10.1016/j.ijfoodmicro.2015.08.015

Leclercq, C., Arcella, D., Piccinelli, R., Sette, S., & Le Donne, C. (2009). The Italian National Food Consumption Survey INRAN-SCAI 2005-06: Main Results: In terms of food consumption. Public Health Nutrition, 12(12), 2504–2532. https://doi.org/10.1017/S1368980009005035

Masciopinto, C., Vurro, M., Lorusso, N., Santoro, D., & Haas, C. N. (2020). Application of QMRA to MAR operations for safe agricultural water reuses in coastal areas. Water Research X, 8, 100062. https://doi.org/10.1016/j.wroa.2020.100062

Minhas, P. S., Saha, J. K., Dotaniya, M. L., Sarkar, A., & Saha, M. (2022). Wastewater irrigation in India: Current status, impacts and response options. Science of the Total Environment, 808, 152001. https://doi.org/10.1016/j.scitotenv.2021.152001

Moazeni, M., Nikaeen, M., Hadi, M., Moghim, S., Mouhebat, L., Hatamzadeh, M., & Hassanzadeh, A. (2017). Estimation of health risks caused by exposure to enteroviruses from agricultural application of wastewater effluents. Water Research, 125, 104–113. https://doi.org/10.1016/j.watres.2017.08.028

Mok, H. F., Barker, S. F., & Hamilton, A. J. (2014). A probabilistic quantitative microbial risk assessment model of norovirus disease burden from wastewater irrigation of vegetables in Shepparton, Australia. Water Research, 54, 347–362. https://doi.org/10.1016/j.watres.2014.01.060

Natural Resources Management Ministerial Council. (2006). Australia Guidelines for Water Recycling: Managing Health and Environmental Risks (Phase 1). National Water Quality Management Strategy, 415.

Ofori, S., Puškáčová, A., Růžičková, I., & Wanner, J. (2021). Treated wastewater reuse for irrigation: Pros and cons. Science of the Total Environment, 760. https://doi.org/10.1016/j.scitotenv.2020.144026

Purnell, S., Halliday, A., Newman, F., Sinclair, C., & Ebdon, J. (2020). Pathogen infection risk to recreational water users, associated with surface waters impacted by de facto and indirect potable reuse activities. Science of the Total Environment, 722, 137799.

https://doi.org/10.1016/j.scitotenv.2020.137799

Rock, C. M., Brassill, N., Dery, J. L., Carr, D., McLain, J. E., Bright, K. R., & Gerba, C. P. (2019). Review of water quality criteria for water reuse and risk-based implications for irrigated produce under the FDA Food Safety Modernization Act, produce safety rule. Environmental Research, 172(August 2018), 616–629. https://doi.org/10.1016/j.envres.2018.12.050

Rönnqvist, M., Mikkelä, A., Tuominen, P., Salo, S., & Maunula, L. (2014). Ultraviolet Light Inactivation of Murine Norovirus and Human Norovirus GII: PCR May Overestimate the Persistence of Noroviruses Even When Combined with Pre-PCR Treatments. Food and Environmental Virology, 6(1), 48–57. https://doi.org/10.1007/s12560-013-9128-y

Simhon, A., Pileggi, V., Flemming, C. A., Lai, G., & Manoharan, M. (2020). Norovirus risk at a golf course irrigated with reclaimed water: Should QMRA doses be adjusted for infectiousness? Water Research, 183, 116121. https://doi.org/10.1016/j.watres.2020.116121

Skovgaard, N. (2004). Risk assessments of Salmonella in eggs and broiler chickens. Interpretative summary. International Journal of Food Microbiology, 91(2), 223. https://doi.org/10.1016/s0168-1605(03)00369-6

Stals, A., Van Coillie, E., & Uyttendaele, M. (2013). Viral genes everywhere: Public health implications of PCR-based testing of foods. Current Opinion in Virology, 3(1), 69–73. https://doi.org/10.1016/j.coviro.2012.11.003

Symonds, E. M., Verbyla, M. E., Lukasik, J. O., Kafle, R. C., Breitbart, M., & Mihelcic, J. R. (2014). A case study of enteric virus removal and insights into the associated risk of water reuse for two wastewater treatment pond systems in Bolivia. Water Research, 65, 257–270. https://doi.org/10.1016/j.watres.2014.07.032

The European Parliament and the Council of the European Union. (2020). Regulation (EU) 2020/741 of 25 May 2020 on minimum requirements for water reuse. Official Journal of the European Union, 2019(April), L 177/32-L 177/55.

Turolla, A., Cattaneo, M., Marazzi, F., Mezzanotte, V., & Antonelli, M. (2018). Antibiotic resistant bacteria in urban sewage: Role of full-scale wastewater treatment plants on environmental spreading. Chemosphere, 191, 761–769. https://doi.org/10.1016/j.chemosphere.2017.10.099

Turolla, Andrea, Sabatino, R., Fontaneto, D., Eckert, E. M., Colinas, N., Corno, G., Citterio, B., Biavasco, F., Antonelli, M., Mauro, A., Mangiaterra, G., & Di Cesare, A. (2017). Defence strategies and antibiotic resistance gene abundance in enterococci under stress by exposure to low doses of peracetic acid. Chemosphere, 185, 480–488. https://doi.org/10.1016/j.chemosphere.2017.07.032

Ungureanu, N., Vlăduț, V., & Voicu, G. (2020). Water scarcity and wastewater reuse in crop irrigation. Sustainability (Switzerland), 12(21), 1–19. https://doi.org/10.3390/su12219055

Van Abel, N., Schoen, M. E., Kissel, J. C., & Meschke, J. S. (2017). Comparison of Risk Predicted by Multiple Norovirus Dose–Response Models and Implications for Quantitative Microbial Risk Assessment. Risk Analysis, 37(2), 245–264. https://doi.org/10.1111/risa.12616

Weir, M. H., Mitchell, J., Flynn, W., & Pope, J. M. (2017). Development of a microbial dose response visualization and modelling application for QMRA modelers and educators. Environmental Modelling and Software, 88, 74–83. https://doi.org/10.1016/j.envsoft.2016.11.011

World Health Organization. (2006). Safe Use of Wastewater , Excreta and Greywater Guidelines for the Safe Use of. World Health, II, 204. http://whqlibdoc.who.int/publications/2006/9241546832_eng.pdf

Zhiteneva, V., Carvajal, G., Shehata, O., Hübner, U., & Drewes, J. E. (2021). Quantitative microbial risk assessment of a non-membrane based indirect potable water reuse system using Bayesian networks. Science of the Total Environment, 780. https://doi.org/10.1016/j.scitotenv.2021.146462