Model-Based Optimisation of Outbreak Detection via Wastewater Probing

Martin Bicher * **Fabian Amman** ** **Gergely Ódor** **
Andreas Bergthaler ** **Niki Popper** ***

* *Institute of Information Systems Engineering, TU Wien, Favoritienstraße 9-11, 1040 Vienna, Austria (e-mail: martin.bicher@tuwien.ac.at)*
** *Center for Pathophysiology, Infectiology and Immunology, Medical University of Vienna, Kinderspitalgasse 15, 1090 Vienna, Austria*
*** *dwh GmbH, Neustiftgasse 57-59, 1070 Vienna, Austria*

1. INTRODUCTION

At least since the COVID-19, crisis wastewater-based epidemiology (WBE) has been recognised internationally as a reliable surveillance and early warning system to track circulating pathogens (Singer et al. (2023)). By probing and sequencing wastewater, WBE allows to detect, quantify and characterise pathogens circulating in the connected catchment. Several health authorities have set up WBE programs in their surveillance efforts. For example in Austria, wastewater from 48 purification plants is probed on a weekly basis for SARS-CoV-2 and its variants (see Amman et al. (2022)), and in parts of Austria also for influenza and the respiratory syncytial virus. Expansion to other relevant pathogens is considered.

The main advantage of WBE over traditional case-based surveillance is its scalability, allowing one sample to cover thousands without requiring active participation, reducing costs and bias. While cost-effective, nationwide surveillance at reasonable granularity incurs significant taxpayer costs. Deciding which wastewater plants to sample and how often is crucial for public health but must be economically justified to convince policymakers and the public of WBE's value.

In this work, we present a model that simulates the regional spread of a new pathogen, its concentration at wastewater plants, and the limit of detection of pathogen specific assays as a function of wastewater catchment characteristics. The goal is to minimise detection time by optimising the selection of plants and sampling intervals. After introducing the model and showing preliminary results with manually varied strategies, we propose ideas for using a simheuristic to solve this optimisation problem.

2. METHODS

2.1 Simulation Model

We follow the network-based SIRS approach in Hethcote (1978) and specify a model with M nodes, representing households, and a corresponding vector $(N_i)_{i=1}^M$ of inhabitants. The nodes are connected by a weighted digraph identified by adjacency matrix $A \in (\mathbb{R}^+)^{M\times M}$, whereas each entry $A_{i,j}$ corresponds to the average daily number of contacts between individuals in nodes i and j. With $S_i(0) + I_i(0) + R_i(0) = N_i$ the dynamics is defined by

$$\begin{aligned} \dot{S}_i(t) &= -S_i(t)\Theta(I,i) &&+\delta R_i(t), \\ \dot{I}_i(t) &= \ S_i(t)\Theta(I,i) &&-\gamma I_i(t), \\ \dot{R}_i(t) &= &&\gamma I_i(t) \ -\delta R_i(t), \end{aligned} \tag{1}$$

$$\Theta(I,i) = A_{i,i}\beta^{in}\frac{I_i(t)}{N_i} + \sum_{j\neq i} A_{i,j}\beta^{out}\frac{I_j(t)}{N_j}. \tag{2}$$

Hereby, β^{in} and β^{out} refer to the in-node and out-node transmission rate, γ to the recovery rate and δ to the immunity waning rate. Time-unit will be days.

Compartments $S_i(t), I_i(t), R_i(t)$ represent the expected number of susceptible, infectious and recovered persons in household i at time t. We argue, that I_i is proportional to the overall quantity of pathogen present in household i and to the pathogen load excreted into the wastewater.

Furthermore we introduce K purification plant nodes and include them into the model using adjacency matrix $B \in \{0,1\}^{K\times M}$. Hereby $B_{k,i} = 1 \Leftrightarrow$ household j lies in the catchment area of plant k. We define

$$W_k(t) = \frac{\sum_{i=1}^M B_{k,i} I_i(t)}{\sum_{i=1}^M B_{k,i} N_i} \tag{3}$$

as the pathogen *ground truth* at plant k. It models the ratio of all pathogen excreted by households in the catchment area compared to a human control signal.

We finally define the *measured signal* at plant k via

$$\hat{W}_k(t) := \nu_k \begin{cases} W_k\left(\left\lfloor \frac{t}{\xi_k} \right\rfloor \xi_k\right), & \text{if } W_k\left(\left\lfloor \frac{t}{\xi_k} \right\rfloor \xi_k\right) > \mu, \\ 0, & \text{else.} \end{cases} \tag{4}$$

Hereby, μ defines a concentration threshold below which a given pathogen signal cannot reliably be detected by a probe. The parameter vectors ν and ξ will be regarded as *probing strategy*: $\nu_k \in \{0,1\}$ defines if plant k is probed at all, $\xi_k \in \mathbb{N}$ defines the interval between taking two probes from the plant.

2.2 Optimisation Problem

For optimisation we focus on the *detection time*, i.e. the first time a signal is detected at any of the probed plants:

⋆ This research was funded in part by the Austrian Science Fund (FWF) I 5908-G.

DOI: 10.34726/9004

$$\tau := \min_{k\in\{1,\dots,K\}} \tau_k^i := \min_{k\in\{1,\dots,K\}} \min_{t>0} \left(\hat{W}_k(t) > 0\right). \quad (5)$$

In addition to the probing strategy, the value of τ will also depend on the location of the outbreak. To simulate the latter we specify the initial conditions $R_i(0) := 0$,

$$I_i(0) := \delta_{i,i_0} I_\epsilon, \text{ and } S_i(0) := N_i(0) - I_i(0), \quad (6)$$

where node i_0 will be regarded as *index-household*. With this definition, (ν, ξ) and i_0 uniquely map to a detection time: $\tau = \tau^{i_0,\nu,\xi}$. Considering that the probing strategy should be able to quickly detect the pathogen independent of the choice of i_0, we specify the optimisation target

$$F : \{0,1\}^K \times \mathbb{N}^K \to \mathbb{R}^+ : F(\nu,\xi) \mapsto \frac{1}{M}\sum_{i_0=1}^{M} \tau^{i_0,\nu,\xi}. \quad (7)$$

Clearly, minimising F alone would be meaningless, since it does not incorporate any cost or sensitivity constrains. We define

$$C_1(\nu,\xi) = \sum_{i=1}^{K} \nu_i, \quad C_2(\nu,\xi) = 7.0 \cdot \sum_{i=1}^{K} \frac{\nu_i}{\xi}. \quad (8)$$

That means, $C_1(\nu,\xi) \leq c_1$ limits the total number of included plants to c_1, to restrict the logistic efforts, and $C_2(\nu,\xi) \leq c_2$ limits the total number of probes taken per week to c_2, to restrict the total costs.

In summary, we define the optimisation problem as follows:

$$(\nu,\xi)_{opt} = \text{argmin}_{(\nu,\xi)\in\Omega_{c_1,c_2}} F(\nu,\xi), \text{ with}$$
$$\Omega_{c_1,c_2} := \{(\nu,\xi) : C_1(\nu,\xi) \leq c_1, C_2(\nu,\xi) \leq c_2\}. \quad (9)$$

3. RESULTS

For the preliminary results shown in this paper, we extracted a synthetic contact network from an existing agent-based epidemics model. It was developed and applied in the course of the COVID-19 crisis, is specified in the SI of Bicher et al. (2021), and was used for export of synthetic data before (see Popper et al. (2021)). We counted, averaged and exported randomly sampled contacts between the roughly 4.5M model households over ten simulated days, leading to an integer vector $N \in \mathbb{N}^{4.5\cdot10^6}$ and a sparse matrix $A \in \mathbb{N}^{4.5\cdot10^6 \times 4.5\cdot10^6}$. Data about the catchment areas of the $K = 636$ largest purification plants in Austria was used to specify B. Finally, disease parameters $\beta^{in}, \beta^{out}, \gamma$ and δ were manually chosen to reflect R_0 and immunity waning behaviour of SARS-CoV-2. To make target function F computeable, we ran the sum defined in (7) only over 5000 randomly drawn index households instead of all M possible ones.

In Figure 1 detection times $\tau^{i,\nu,\xi}$ are compared for two probing strategies. The *reference* strategy uses the 48 Austrian plants probed once per week, analogous to the currently implemented system in Austria. The *comparator* strategy uses 27 manually selected plants probed in intervals between 2 and 12 days. Both lie in $\Omega_{48,48}$ and are therefore comparable with respect to costs.

4. DISCUSSION AND CONCLUSION

Preliminary results show that the currently implemented strategy for early pathogen detection can be improved. Manual tests reduced detection time by nearly three days,

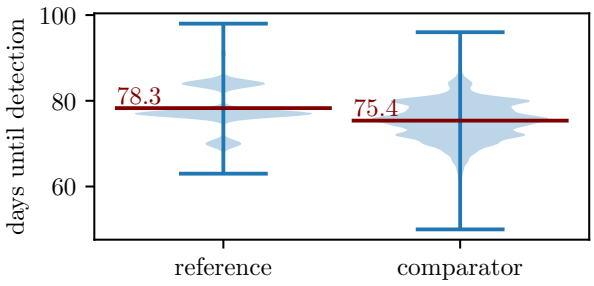

Fig. 1. Detection times for two probing strategies in $\Omega_{48,48}$. The red line shows the target F computed as the mean of the detection-times for 5000 random index households.

offering policymakers a crucial time-advantage. Using well-defined simheuristics instead of manual methods would yield further improvements.

The optimisation challenge lies in the time-consuming simulations and vast search space. Despite efficient matrix multiplications and parallelisation using Numpy and Scipy, computing I, W, and $\hat{W}$ for one index household and probing strategy still takes about one minute on a well-equipped server.

Given the complexity of the search space Ω_{c_1,c_2}, traditional population-based metaheuristics like Genetic Algorithms (GAs) would require large populations to converge effectively. A more integrated approach is needed to reduce the number of simulations required. The plan is to exploit additional feedback from the simulation in addition to the detection time, such as particularly successful or unsuccessful plants, to guide more targeted crossovers and mutations in a GA setup.

REFERENCES

Amman, F., Markt, R., Endler, L., Hupfauf, S., Agerer, B., Schedl, A., Richter, L., Zechmeister, M., Bicher, M., Heiler, G., et al. (2022). Viral variant-resolved wastewater surveillance of sars-cov-2 at national scale. *Nature biotechnology*, 40(12), 1814–1822.

Bicher, M., Rippinger, C., Urach, C., Brunmeir, D., Siebert, U., and Popper, N. (2021). Evaluation of Contact-Tracing Policies against the Spread of SARS-CoV-2 in Austria: An Agent-Based Simulation. *Medical Decision Making*, 41(8), 1017–1032. doi: 10.1177/0272989X211013306.

Hethcote, H.W. (1978). An immunization model for a heterogeneous population. *Theoretical population biology*, 14(3), 338–349.

Popper, N., Zechmeister, M., Brunmeir, D., Rippinger, C., Weibrecht, N., Urach, C., Bicher, M., Schneckenreither, G., and Rauber, A. (2021). Synthetic reproduction and augmentation of covid-19 case reporting data by agent-based simulation. *Data Science Journal*, 20, 16. doi: 10.5334/dsj-2021-016.

Singer, A.C., Thompson, J.R., Filho, C.R.M., Street, R., Li, X., Castiglioni, S., and Thomas, K.V. (2023). A world of wastewater-based epidemiology. *Nature Water*, 1(5), 408–415.