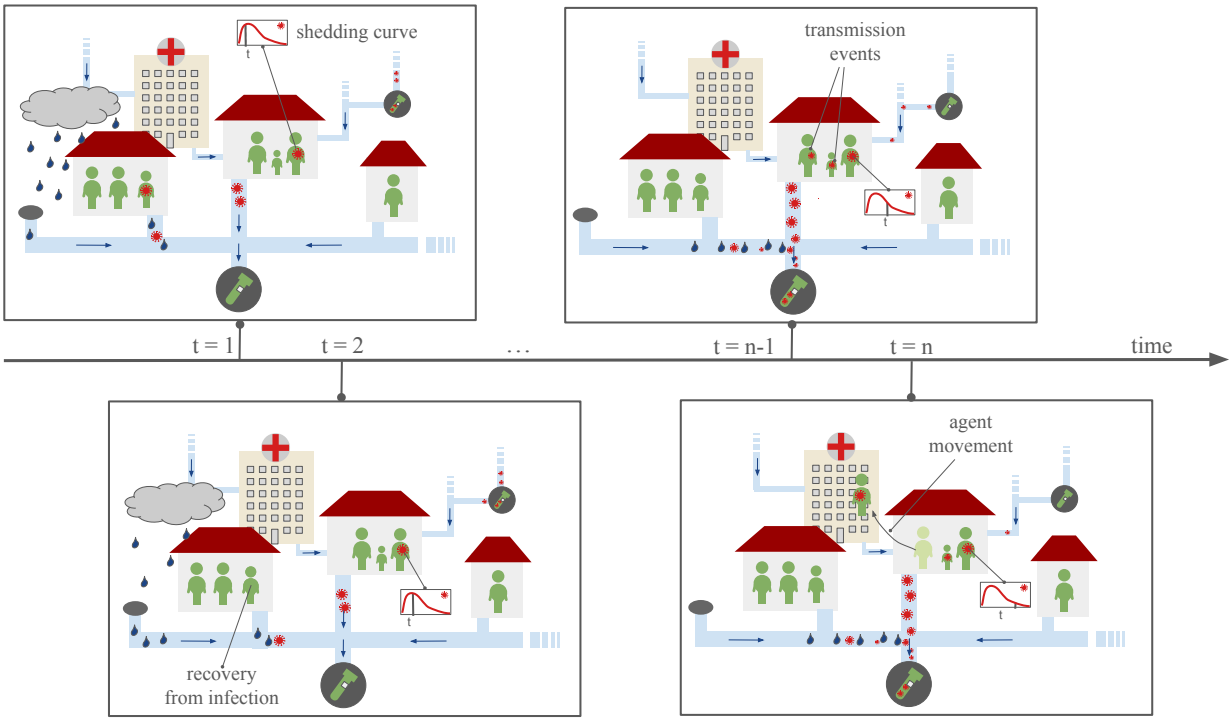
Graphical Abstract

Integrative Modeling of the Spread of Serious Infectious Diseases and Corresponding Wastewater Dynamics

Nina Schmid, Julia Bicker, Andreas F. Hofmann, Karina Wallrafen-Sam, David Kerkmann, Andreas Wieser, Martin J. Kühn, Jan Hasenauer



Highlights

Integrative Modeling of the Spread of Serious Infectious Diseases and Corresponding Wastewater Dynamics

Nina Schmid, Julia Bicker, Andreas F. Hofmann, Karina Wallrafen-Sam, David Kerkmann, Andreas Wieser, Martin J. Kühn, Jan Hasenauer

- Integration of infection and wastewater models to simulate disease spread.
- Identification of factors affecting wastewater measurements.
- Illustration of ability of wastewater-based surveillance to predict outbreaks before case reporting.
- Demonstration of unreliability of flow rate normalization in case of rainwater infiltration.
- Optimization of wastewater-based surveillance for improved public health monitoring.

Integrative Modeling of the Spread of Serious Infectious Diseases and Corresponding Wastewater Dynamics

Nina Schmid[a], Julia Bicker[b], Andreas F. Hofmann[c], Karina Wallrafen-Sam[a], David Kerkmann[d], Andreas Wieser[e,f,g], Martin J. Kühn[a,b], Jan Hasenauer[a,h,i]

[a] *Life & Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany*

[b] *Institute of Software Technology, Department of High-Performance Computing, German Aerospace Center, Cologne, Germany*

[c] *tandler.com GmbH, Buch am Erlbach, Germany*

[d] *Helmholtz Centre for Infection Research, Brunswick, Germany*

[e] *Institute of Infectious Diseases and Tropical Medicine, LMU University Hospital Munich, Munich, Germany*

[f] *German Centre for Infection Research (DZIF), Partner Site Munich, Munich, Germany*

[g] *Max von Pettenkofer Institute, Faculty of Medicine, LMU Munich, Munich, Germany*

[h] *Technische Universität München, Center for Mathematics, Garching, Germany*

[i] *Helmholtz Zentrum München, German Research Center for Environmental Health, Computational Health Center, Neuherberg, Germany*

Abstract

The COVID-19 pandemic has emphasized the critical need for accurate disease modeling to inform public health interventions. Traditional reliance on confirmed infection data is often hindered by reporting delays and under-reporting, while widespread antigen and antibody testing can be costly and impractical. Wastewater-based surveillance offers a promising alternative by detecting viral concentrations from fecal shedding, potentially providing a more accurate estimate of true infection prevalence. However, challenges remain in optimizing sampling protocols, locations, and normalization strategies, particularly in accounting for environmental factors like precipitation.

We present an integrative model that simulates the spread of serious infectious diseases by linking detailed infection dynamics with wastewater processes through viral shedding curves. Through comprehensive simulations, we examine how virus characteristics, precipitation events, measurement protocols, and normalization strategies affect the relationship between infection dynamics and wastewater measurements. Our findings reveal a complex relationship between disease prevalence and corresponding wastewater concentrations, with

Email address: nina.schmid@uni-bonn.de (Nina Schmid)

key variability sources including upstream sampling locations, continuous rainfall, and rapid viral decay. Notably, we find that flow rate normalization can be unreliable when rainwater infiltrates sewer systems. Despite these challenges, our study demonstrates that wastewater-based surveillance data can serve as a leading indicator of disease prevalence, predicting outbreak peaks before they occur. The proposed integrative model can thus be used to optimize wastewater-based surveillance, enhancing its utility for public health monitoring.

Keywords: infectious diseases, wastewater, agent-based model, hydrodynamic model, sewer network, shedding model

---

1. Introduction

The COVID-19 pandemic has highlighted the need for effective real-time monitoring and prediction of infectious disease dynamics to support timely and informed intervention policies. In this context, various studies have been conducted to assess vaccine distribution strategies [6, 12] or the effectiveness of non-pharmaceutical interventions like telework suggestions, prohibition of private gatherings of certain sizes, or partial lock-downs [39, 7, 27, 19].

The spread of infectious diseases is nowadays modeled using a broad range of approaches, including statistical and machine learning models [39, 27, 33], compartmental and meta-population models [10, 6, 45], and agent-based models [7, 29, 22, 15], or even hybrid approaches [18, 4]. Among these approaches, agent-based models (ABMs) allow for the most detailed description of disease dynamics. ABMs simulate the spread of infectious diseases on an individual level, thereby facilitating the incorporation of comprehensive information about localization, interaction, and behavior. The models are intrinsically stochastic and based on discrete- or continuous-time Markov processes. While ABMs are the state-of-the-art in infectious disease modeling, their advancement remains an active field of research. A key challenge is the choice of model parameters.

The predictive power of models for the spread of infectious diseases depends on the available data and the ability to incorporate them into models. The number of confirmed infections is the most common data source. However, confirmed infections are affected by reporting delays [26] and subject to under-reporting [28], limiting the reliability of the resulting models [35]. While single antigen and antibody tests can be cheap, the use of these tests to study large, representative population cohorts is resource intensive. Furthermore, even the cohort might be subject to sampling bias [30, 17]. Wastewater-based surveillance presents a promising solution to these issues by capturing viral concentrations from all infected individuals within a catchment area, including those who are asymptomatic or undetected by traditional testing methods. Via detection of viral RNA in sewage, outbreaks can be identified before clinical cases are reported. This capability was demonstrated during the COVID-19 pandemic, when several national wastewater-based surveillance programs were

established, e.g. the AMELAG project in Germany [37]. Wastewater-based surveillance is applicable to various diseases detectable in sewage, including poliovirus, hepatitis, norovirus, and influenza [23], but also to monitoring antimicrobial resistance [9]. Yet, infectious disease monitoring based on wastewater-based surveillance data still presents several significant challenges: The choice of the sampling location can yield different viral concentrations due to the interplay of population density and sewer infrastructure. The impact of viral degradation and variations in flow-time on measurements is not well understood. The environmental conditions, e.g. rain fall, can impact measurement and the effectiveness of established normalization strategies, such as flow-based adjustments, remains unclear. Despite these challenges, integrating wastewater data with traditional infectious disease models holds promise for improving prediction accuracy and has been attempted in several studies [8, 32].

The challenges of wastewater-based surveillance data can in principle be addressed using comprehensive computational models which provide in-depth descriptions of the spread of infectious diseases as well as wastewater dynamics. Yet, to the best of our knowledge, most published studies use relatively simple approaches. For example, Wu et al. [43] calculated a rough estimate of SARS-CoV-2 prevalence upstream of a wastewater treatment facility using the normalized viral load measured in twelve wastewater samples and assumptions about the sewer system flow volume, stool sizes, and average viral concentration in stool among infected persons. The authors concluded that prevalence in their population of interest was much higher than the confirmed case count, even under conservative assumptions, but noted that their estimate was subject to considerable uncertainty, as they did not account for the timeline of viral shedding or the loss of viral copies along sewer lines, among other factors. Hart and Halden [14] used a simplified hydrodynamic model of a city sewer network to estimate SARS-CoV-2 detectability in wastewater under different temperature-driven decay scenarios. This study highlighted the importance of appropriately accounting for viral decay when analyzing wastewater data, but the authors assumed exclusively dry weather conditions and their consideration of the relationship between SARS-CoV-2 prevalence and viral load entering the sewer system was limited; accounting for variations in viral shedding across individuals and over time was out of the study's scope. Peccia et al. [34] compared wastewater-based surveillance data to positive COVID-19 tests and hospital admissions using a basic distributed lag time series model and found that the former led the latter data by several days. This approach highlighted the potential of wastewater data to provide early warnings of outbreaks, but relied on the assumption that the observed wastewater measurements were unbiased. Finally, Nourbakhsh et al. [32] coupled an SEIR-type compartment model of SARS-CoV-2 transmission with a simple advection-dispersion-decay model of virus concentration dynamics in a sewer system to estimate cumulative incidence in several cities based on empirical wastewater measurements and reported case data. However, because of a limited sewer system model and the use of ordinary differential equations, the model was not well adapted for scenarios such as small communities or low-prevalence settings. To the best

of our knowledge, no existing study on wastewater-based infectious disease monitoring uses advance methods from the well-established field of sewage network modeling [38]. This is problematic as wastewater dynamics in sewage networks are highly complex and require advanced modeling tools for the prediction of pollutant load [2]. Accordingly, advances beyond these existing methodological frameworks could significantly enhance our understanding of disease spread and lead to more effective public health interventions.

Our work contributes to the field by integrating an agent-based model (ABM) for infection dynamics, a viral shedding model, and a detailed hydrodynamic model of sewage flow and viral load. We provide the mathematical details and a numerical implementation. Through simulation studies for a respiratory virus (Section 3.1), we investigate the effects of measurement protocols (Section 3.3 and Section 3.4), precipitation events (Section 3.5), viral decay (Section 3.6), and normalization strategies (Section 3.7) on the relationship between infection dynamics and wastewater measurements. This controlled setting enables an in-depth model-based analysis and practical recommendations for real-world wastewater-based surveillance. Our findings advance the development of detailed integrative models informed by data, enhancing the accuracy and reliability of infectious disease monitoring and prediction. Furthermore, the integrated model provides a basis for coherent data integration.

2. Mathematical Model

To study wastewater-based surveillance data, we combine state-of-the-art models for the spread of infectious diseases and wastewater dynamics (Fig. 1). The link is established using a viral shedding model. In this section, we outline the mathematical formulations of the individual models' components and their simulation algorithms as well as their integration. All models are dynamic and are executed on the same time axis such that a coupling with comparisons of outputs is possible.

2.1. Overview

The proposed model consists of three modules:

The **infection dynamics model** describes the time-dependent location and infection state of individual persons, in the following also denoted as agents. The infection state of agents can change due to events such as virus transmission, worsening of symptoms, or recovery. The likelihood of an agent infecting others is determined by its viral load, which varies throughout its infection course, and the length of contact.

The **shedding model** describes the release of virus and viral fragments from infected individuals into their surroundings. It is used in the infection dynamics model to determine the transmission probability as well as to describe the release through urine and stool. The

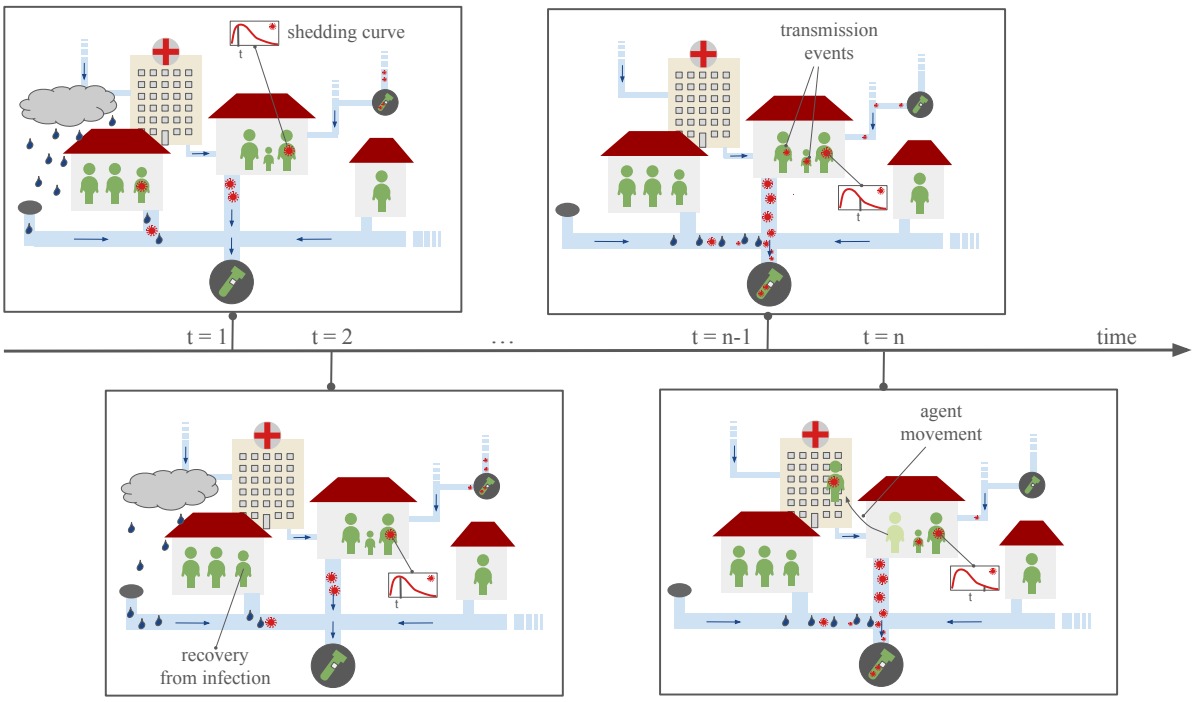


Figure 1: **Model Visualization.** Based on infection dynamics and the movement of agents, RNA is shed at specific locations and time points into the sewer system. This information is propagated over time to simulate the concentration measurements in wastewater. Among other elements, the model can account for precipitation events, sampling protocols, viral load dynamics, and RNA degradation. The figure only depicts a simplified neighborhood with locations of type home or hospital, while the model used for this study allows for various types of agent movements (Section 2.2.1).

shedding curve is assumed to be dependent on the individual's viral load, which is time-dependent and initially increases before declining as the host's immune response takes effect. The shedding model uses the infection state of agents and their time since transmission, which are provided by the infection dynamics model, to determine the viral RNA entering the wastewater system.

The **wastewater dynamics model** simulates the transport and degradation of viral RNA within the sewage network. Using the viral shedding input from the shedding model as well as the agents' locations from the infection dynamics model, this module calculates the RNA concentrations at various points in the network, accounting for factors such as viral decay, flow rates, and the architecture of the sewer system. The output of this model is the RNA concentration at different sampling points, which provides a comprehensive picture of the data to be expected from wastewater-based surveillance.

In the current model, the wastewater does not influence the infection dynamics; accordingly, the integrated model possesses a hierarchical structure. In the following, we discuss the individual modules in more detail.

2.2. Modeling Infection Dynamics

In this study, we use an *agent-based model* (ABM) implemented in the software framework MEmilio [24] to simulate disease states and mobility patterns at the *agent* level, providing a fine-grained view of disease dynamics. It comprises agents with different attributes.

The properties of an *agent* α are defined via an m-tuple $(a_1, ..., a_m) \in \Omega$ with m different attributes a_i, $i = 1, ..., m$. The attributes can be static – meaning that they do not change over the course of the simulation – or dynamic. The static attributes are:

- An agent's age group $\mathcal{A}^{(\alpha)} \in \{1, ..., n_A\}$, with n_A denoting the total number of age groups.
- An agent's set of locations $\mathcal{L}^{(\alpha)} = \{l_j\}_{j \in K^{(\alpha)}}$ with $K^{(\alpha)} \subset \{1, ..., n_L\}$ denoting the subset of all locations it can theoretically move to during a simulation.

The dynamic attributes are:

- An agent's current location $l^{(\alpha)} \in \mathcal{L}^{(\alpha)}$.
- An agent's current infection state $s^{(\alpha)} \in \mathcal{S}$, with $\mathcal{S}$ denoting a set of infection states.
- An agent's time since virus exposure $\tau^{(\alpha)}$ in hours, which is set to NaN if the agent has not been infected.

The simulation of the ABM provides information about the agent's trajectory in space and infection state. In the following, we provide additional details on the ABM, providing the basis for the simulation of mobility (Algorithm 1) and the full population dynamics (Algorithm 2).

2.2.1. Mobility Model

The ABM uses a location graph with n_L locations, $l_1, ..., l_{n_L}$, to model mobility. Every location has a *location type* $T \in \mathcal{T}$ with $\mathcal{T}$ being a set of location types such as *Home*, *School*, *Work*, *Recreation*, *Shop*, *Hospital*, and *Intensive Care Unit (ICU)*. There can be multiple locations of the same type. In addition to the type, a location also has a capacity specifying the maximum number of agents that can enter the location and a maximum number of contacts an agent can have at the location. Every agent α has a set of n_α many locations $\mathcal{L}^{(\alpha)} = \{l_j\}_{j \in K^{(\alpha)}}$, with $K^{(\alpha)} \subset \{1, ..., n_L\}$ being an index set, that are assigned to it. These assigned locations are the ones the agent can move between, meaning that it is restricted to a subgraph of the global location graph (see Supplementary Fig. B.1(b)). We denote the current location of an agent α at a given time point $\tilde{t}$, with $\tilde{t}$ given in hours, as $l^{(\alpha)}(\tilde{t}) \in \mathcal{L}^{(\alpha)}$. Movements, i.e. location transitions, are modeled by an ordered set of mobility

Algorithm 1: ABM Mobility.

Input: Agent α, time point $\tilde{t}_k$, agent's location $l^{(\alpha)}(\tilde{t}_k)$
Output: Agent's location $l^{(\alpha)}(\tilde{t}_{k+1})$
Set $l_{new} = l^{(\alpha)}(\tilde{t}_k)$
Forall mobility rules $m_i \in \mathcal{M}$
  If $m_i(\alpha, \tilde{t}_k) \neq T(l_{new})$
    Get location $l \in \mathcal{L}^{(\alpha)}$ such that $T(l) = T_{to}^{(i)}$
    If *capacity*(l) not reached
      $l_{new} = l$
      return l_{new}

rules $\mathcal{M} = \{m_1, ..., m_{n_M}\}$ that have probabilistic components, which cause stochasticity between simulations. These mobility rules include daily regular behavior like going to work or school on weekdays, irregular behavior like occasionally attending a social event, and behavior related to the infection state of an agent, e.g. going to hospital when having severe symptoms. For Ω containing all potential states of any agent and $[\tilde{t}_0, \tilde{t}_{max}]$ denoting the simulation period, a mobility rule $m_i : \Omega \times [\tilde{t}_0, \tilde{t}_{max}] \to \mathcal{T}$ is given by

$$m_i(\alpha, \tilde{t}) = \begin{cases} T_{to}^{(i)}, & \text{if } \delta^{(i)}(\tilde{t}, \alpha)\ \delta_{T_{from}^{(i)}}\left[T(l^{(\alpha)}(\tilde{t}))\right] X_i = 1 \\ T(l^{(\alpha)}(\tilde{t})), & \text{else} \end{cases} \tag{1}$$

with $T(l) \in \mathcal{T}$ denoting the type of location l, $X_i \in \{0, 1\}$ denoting a Bernoulli distributed random variable with probability p_i, which has a different value for every mobility rule, and $\delta_*[\cdot]$ denoting binary-valued functions defined as

$$\delta_{T_{from}^{(i)}}\left[T(l^{(\alpha)})\right] = \begin{cases} 1, & \text{if } T(l^{(\alpha)}) = T_{from}^{(i)}, \\ 0, & \text{else.} \end{cases} \tag{2}$$

Furthermore, $T_{from}^{(i)}, T_{to}^{(i)} \in \mathcal{T}$ are location types and $\delta^{(i)}(\tilde{t}, \alpha)$ are binary-valued functions provided in Appendix A together with p_i for each mobility rule. The location $l^{(\alpha)}$ of agent α at time point $\tilde{t}$, with $\tilde{t}$ in hours, is given by $l^{(\alpha)}(\tilde{t}) = f(\alpha, \tilde{t})$ (see Algorithm 1), which is evaluated at discrete time points $\tilde{t}_0, ..., \tilde{t}_{max}$ given a time step $\Delta\tilde{t}$ and $\tilde{t}_{k+1} = t_k + \Delta\tilde{t}$.

2.2.2. Disease Progression

An agent α has a time-dependent infection state $s^{(\alpha)}(t)$ from a set of infection states $\mathcal{S}$. The infection states used for this study are *Susceptible* (S), *Exposed* (E), *Non-symptomatically*

Infected (I_{ns}), *Symptomatically Infected* (I_{sy}), *Severely Infected* (I_{sev}), *Critically Infected* (I_{cri}), *Recovered* (R), and *Dead* (D). Infection state I_{ns} includes infectious pre- and asymptomatic agents and infection state I_{sev} includes agents requiring hospital treatment while I_{cri} includes agents requiring ICU treatment. We call an agent *infected* if it has infection state $s^{(\alpha)}(t) \in \{E, I_{ns}, I_{sy}, I_{sev}, I_{cri}\}$ and *formerly infected* if $s^{(\alpha)}(t) \in \{R, D\}$. Transitions between infection states are stochastic and possible either through virus transmission ($S \to E$) (see Section 2.2.3) or disease progression ($E \to I_*, I_* \to I_{**}, I_{**} \to \{R, D\}$) (see Supplementary Fig. B.1(a)). For a (formerly) infected agent α, the course of infection is defined as $\mathcal{I}^{(\alpha)} = \{(t_1^{(\alpha)}, s_1^{(\alpha)}), \ldots, (t_{h_\alpha}^{(\alpha)}, s_{h_\alpha}^{(\alpha)})\}$ containing the time points $t_1^{(\alpha)}, \ldots, t_{h_\alpha}^{(\alpha)}$ at which the agent changes or changed its infection state and the corresponding infection states $s_1^{(\alpha)}, ..., s_{h_\alpha}^{(\alpha)}$. Hence, $t_1^{(\alpha)}$ is the time point at which the agent is exposed and $t_{h_\alpha}^{(\alpha)}$ the time point at which the agent recovers or dies, i.e. $s_{h_\alpha} \in \{R, D\}$. The intermediate time values $t_2^{(\alpha)}, ..., t_{h_\alpha - 1}^{(\alpha)}$ are the time points at which an agent changes to one of the infectious states of the agent's individual course. The length h_α of the course of infection differs between non-symptomatic and (severe or critical) symptomatic courses and is therefore agent-dependent. The stay times in infection states $E, \ldots, I_{cri}$ are log-normally distributed and the transitions $E \to I_*, I_* \to I_{**}, I_{**} \to \{R, D\}$ between all infection states, apart from virus transmission ($S \to E$), are Bernoulli distributed; see Supplementary Table C.1 for the values of all disease progression-related parameters used in the results section.

2.2.3. Disease Transmission

Infected agents can transmit the virus to susceptible agents if they are at the same location. For a susceptible agent in location l, the waiting time until transmission is exponentially distributed with rate $\Lambda_l(t)$. Assuming that agents change their locations only at discrete time points, $t_0, ..., t_{max}$, the number of agents at a location is constant in $[t_k; t_{k+1})$ and $\Lambda_l(t)$ is given by

$$\Lambda_l(t) = \sum_{\alpha \in l} \lambda_\alpha(t). \tag{3}$$

for $t \in [t_k; t_{k+1})$ given in days. We use the casual notation $\alpha \in l$ to iterate over the infected agents at location l in the interval $[t_k; t_{k+1})$. The agent-dependent rate $\lambda_\alpha(t)$ is given by the infectiousness curve described in Section 2.3, Eq. (8). The location-specific infection rate (3) builds on the assumption of homogeneous mixing within the location. If the waiting time until transmission is longer than the time until the susceptible agent leaves location l, no transmission occurs. Hence, for a fixed time step $\Delta t = t_{k+1} - t_k$ given in days, the probability that a susceptible agent at location l gets exposed is

Algorithm 2: ABM Simulation.

Input: Start time t_0, end time t_{max}, time step Δt in days, parameters as described above
Initialize locations
  Create desired number of locations for every type
  Set capacity and maximum number of contacts
Initialize agents
  Set age group
  Assign locations and set initial location
  Set initial infection state and sample course of infection for initially infected
Set $t = t_0$
While $t < t_{max}$
  Forall agents α
    If $s^{(\alpha)} = S$
      Calculate $\Lambda_{l^{(\alpha)}}(t)$
      Draw waiting time $v \sim Exp(\Lambda_{l^{(\alpha)}}(t + \frac{\Delta t}{2}))$
      If $v \leq \Delta t$
        Sample course of infection $\mathcal{I}^{(\alpha)}$
      Else
        continue
    Forall mobility rules m_i
      If $T(l^{(\alpha)}) \neq m_i(\alpha, 24 \cdot t)$ **and** capacity$(f(\alpha, 24 \cdot t))$ not reached
        $l^{(\alpha)} = f(\alpha, 24 \cdot t)$
        break
  $t = t + \Delta t$

$$1 - e^{-\Delta t \int_{t_k}^{t_{k+1}} \Lambda_l(t) dt}. \tag{4}$$

2.3. Modeling Viral Load, Infectiousness and Shedding

The viral load of individual patients determines their infectiousness and shedding. Here, we model viral shedding using established models [20, 21]. The viral load for an agent α in RNA copies per swab on the log_{10} scale at a given time t in days (see Supplementary Fig. B.2, left) is defined as

$$v_\alpha(t) = \begin{cases} \frac{v_{\max}^{sy}}{\tau_E + \tau_{I_{ns}}} \cdot (t - t_E), & \text{if } t \in [t_E, t_{v_{\max}}] \\ v_{\max} + -\frac{v_{\max}}{t_{R/D} - t_{v_{\max}}} \cdot (t - t_{v_{\max}}), & \text{if } t \in (t_{v_{\max}}, t_{R/D}] \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

with t_E, $t_{v_{\max}}$, and $t_{R/D}$ denoting the times in days of virus exposure, maximal viral load, and recovery/death, respectively. The time points are agent specific, with $t_E = t_1^{(\alpha)} \leq t_{v_{\max}} \leq t_{R/D} = t_{h_\alpha}^{(\alpha)}$. As the infection state trajectories are sampled at the time point of exposure, $t_{R/D}$ is readily available.

We assume that symptomatically infected agents reach the peak viral load the moment they show symptoms, while agents not showing symptoms reach the peak in the middle of their infection period. Furthermore, we assume agents that share the same duration $\tau_E + \tau_{I_{ns}}$ also share the same linear increase until $t_{v_{\max}}$. This yields for the variables $t_{v_{\max}}$ and $v_{\max}$:

$$t_{v_{\max}} = \begin{cases} t_E + \tau_E + \tau_{I_{ns}}, & \text{if } I_{sy} \in \{s_1^{(\alpha)}, ..., s_{h_\alpha}^{(\alpha)}\} \\ t_E + 0.5 \cdot (t_{R/D} - t_E), & \text{else} \end{cases}$$

$$v_{\max} = \begin{cases} v_{\max}^{sy}, & \text{if } I_{sy} \in \{s_1^{(\alpha)}, ..., s_{h_\alpha}^{(\alpha)}\} \\ \frac{v_{\max}^{sy}}{\tau_E + \tau_{I_{ns}}} \cdot (t_{v_{\max}} - t_E), & \text{else} \end{cases}$$

where $v_{\max}^{sy}$ is the peak viral load given in log_{10} RNA copies per swab for symptomatic infections (see Supplementary Table C.2 for the value of this and all other parameters relevant to viral shedding).

Following the observation of Jones et al. [20], we model the shape of an agent α's shedding by a sigmoid function of their viral load, i.e.,

$$\zeta_\alpha(t) = \frac{1}{1 + \exp\left(-(a + b \cdot v_\alpha(t))\right)} \tag{6}$$

with shape parameters $a, b > 0$. As in [20], since there is no information about when shedding starts after exposure, we make the assumption that shedding is zero or close to it as long as the agent is still in the Exposed state. Therefore we introduce a time shift of $\tau_{shift} = 0.6 \cdot \tau_E$. Thus, the scaled and shifted shedding curve for an agent α is given by:

$$\gamma_\alpha(t) = \begin{cases} \kappa_\gamma \cdot \zeta_\alpha(t - \tau_{shift}), & \text{if } t \in [t_E + \tau_{shift}, t_{R/D}] \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where κ_γ is a scaling factor that translates $\zeta_\alpha(t - \tau_{shift})$ into an RNA shedding rate. The RNA shedding into the sewage network is the number of RNA copies shed in total per day and has to be normalized by the water flushed into the wastewater system to receive a unit of copies per liter.

Similarly, the corresponding unitless infectiousness curve for agent α at time point t (given in days) (see Supplementary Fig. B.2, right) is given by

$$\lambda_\alpha(t) = \begin{cases} \kappa_\lambda \cdot \zeta_\alpha(t - \tau_{shift}), & \text{if } t \in [t_E + \tau_{shift}, t_{R/D}] \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

with κ_λ translating $\zeta_\alpha(t - \tau_{shift})$ into a transmission rate (see Section 2.2.3).

2.4. Modeling the Sewage System and its Hydrodynamics

We use a comprehensive model of a wastewater network to simulate the sewage flow and reactive transport of dissolved chemical substances in wastewater while avoiding simplifications and incorrect interpretations. In this model, a sewer system is represented as a directed acyclic graph defined by n edges $\mathcal{E} = \{e_1, ..., e_n\}$ and m nodes $\mathcal{N} = \{n_1, ..., n_m\}$. The edges represent sewage pipes and the nodes represent junctions as well as entry and exit points. Edges and nodes are characterized by several parameters, including total volume and height. The state of the system is the amount of water contained in the edges and nodes, its flow rate, and the concentrations of the relevant substances. The simulation of water inflow to the system is based on two phenomena: hydraulic surface runoff during precipitation events and water usage of industry and citizens.

At their core, the hydrodynamic calculations are based on the Saint-Venant Equation [16]. The equation assumes one-dimensional flows in open channels and mass and momentum conservation, which yields

$$\frac{1}{g}\frac{\partial v}{\partial t} + \frac{v}{g}\frac{\partial v}{\partial x} + \frac{v}{g \cdot A}q + \frac{\partial h}{\partial x} + \frac{\lambda}{4 \cdot R}\frac{v|v|}{2g} = 0, \tag{9}$$

with flow velocity v in $\frac{\mathrm{m}}{\mathrm{s}}$, time t in s, gravitational acceleration g in $\frac{\mathrm{m}}{\mathrm{s}^2}$, vertical position along the pipe x in m, cross-sectional flow area A in m^2, lateral inflow corresponding to the precipitation of a specified time interval q in $\frac{\mathrm{m}^2}{\mathrm{s}}$, water levels h in m, pipe friction coefficient λ (unitless) and hydraulic radius R in m (i.e. $\frac{A}{\text{Circumference}}$).

Replacing the differential operators with differential quotients for specified time points $t_1 < t_2$

and locations $x_1 < x_2$ yields

$$\frac{\partial v}{\partial t} \to \frac{v(t_2, x) - v(t_1, x)}{t_2 - t_1}, \tag{10}$$

$$\frac{\partial v}{\partial x} \to \frac{v(t, x_2) - v(t, x_1)}{x_2 - x_1}, \tag{11}$$

$$\frac{\partial h}{\partial x} \to \frac{h(t, x_2) - h(t, x_1)}{x_2 - x_1}, \tag{12}$$

for $t \in [t_1, t_2]$ and $x \in [x_1, x_2]$. By interpreting $\Delta x = x_2 - x_1$ as the length of an edge of the wastewater network, Eq. (9) after rearrangement becomes a quadratic equation of the form

$$a v(t, x)^2 + b v(t, x) + c = 0, \tag{13}$$

where

$$\begin{aligned} a &= \text{sign}(v(t_1, x)) \cdot \frac{\lambda}{8R}, \\ b &= \frac{1}{t_2 - t_1} + \frac{v(t, x_2) - v(t, x_1)}{\Delta x} + \frac{\text{q}}{\text{A}} \quad \text{and} \\ c &= -\frac{v(t_1, x)}{t_2 - t_1} + \frac{h(t, x_2) - h(t, x_1)}{\Delta x}\ g\ . \end{aligned}$$

The quadratic equation has two complex solutions, where the flow velocity v is equal to the real part of those solutions:

$$v(t, x) = \text{RE}\left(-b + \frac{\sqrt{b^2 - 4ac}}{2a}\right). \tag{14}$$

Using this solution and the boundary conditions $v(t, x_1)$, $v(t, x_2)$, $h(t, x_1)$, and $h(t, x_2)$, as well as the initial condition $v(t_1, x)$, a solution of the flow velocity can be calculated for arbitrary time points and edges of the network system. The state of the nodes of the system define the boundary and initial conditions. Based on the solutions from Eq. (9) for edges connected to a node, the in- and outflow to the node is given for each edge. At the center of the node, the sum of in- and outflow equals zero (mass conservation). Based on this assumption, the flow rates and heights at the edge borders can be calculated.

While Δx can be defined based on the length of an edge, $\Delta t = t_2 - t_1$ has to be chosen carefully. A too large value of Δt will yield inaccurate numerical solutions, while a too small value will yield unnecessarily high computation times. A complex solution of Eq. (13) with an imaginary component larger than zero indicates the transition to an oscillatory state. To ensure a meaningful numerical solution, Δt is chosen to be smaller than the corresponding oscillation period. Further considerations like the maximum total change of volume provide an equation to set Δt for each iteration of the numerical solution scheme, such that stable solutions of

v are ensured even for extreme hydraulic scenarios (see [41]). By combining Eq. (13) with other known physical principles, e.g. energy loss along a pipe (using the Prandtl-Colebrook equation) and energy loss due to inelastic collisions in manhole structures (according to the Borda-Carnot equation), the simulation precision is further refined.

Based on the calculated flow velocities and other time-varying edge state variables, the substance concentration per location and time point can be calculated. Viral loads generated by agents of the ABM enter the sewer system as concentrations in the respective amount of domestic wastewater generated for every time step of the model simulation, at the location the agent currently occupies. Viral fragments are then transported through the system according to the pre-calculated flow rates, potentially taking chemical reactions in the form of viral decay into account. The outputs of the hydraulic simulation are time- and location-dependent concentration curves. The time step for the viral load calculation is chosen prior to the calculations and a suitable value depends on the specifics of the viral decay dynamics, where faster changing dynamics suggest choosing a smaller time step.

After defining the wastewater network, its connected surfaces and corresponding runoffs, as well as substance characteristics, the simulation proceeds in two steps (see Supplementary Fig. B.3). First, the flow velocities and volumes are calculated with a numerical solver. Secondly, the viral load over time is simulated.

For the numerical simulation, we use the urban water management modeling and simulation environment ++SYSTEMS, developed by the company tandler.com GmbH. Utilizing the mathematical principles described above, ++SYSTEMS with its backend and calculation kernel DYNA forms a fully dynamic, geospatial modeling and management software for waste- and rainwater (individually or combined) sewer systems. Details on the implementation are available in [41].

3. Results

3.1. Demonstrator Setup

To address open questions and challenges related to the interpretation of wastewater-baser surveillance data, we performed a simulation study, which allows us to evaluate counterfactual scenarios without missing data or data uncertainty. To this end, we developed a synthetic, yet realistic model of a city neighborhood and corresponding sewer system (Fig. 2); we then traced infectious disease outbreaks in this controlled setting using our sequence of three modules, from the MEmilio-based infection dynamics model to the shedding model to the ++SYSTEMS-based wastewater dynamics model. The synthetic neighborhood used in our study contains residential, recreational, university, mixed shopping and business, and mixed residential and industry surface areas. It is populated by at least 838 agents, whose

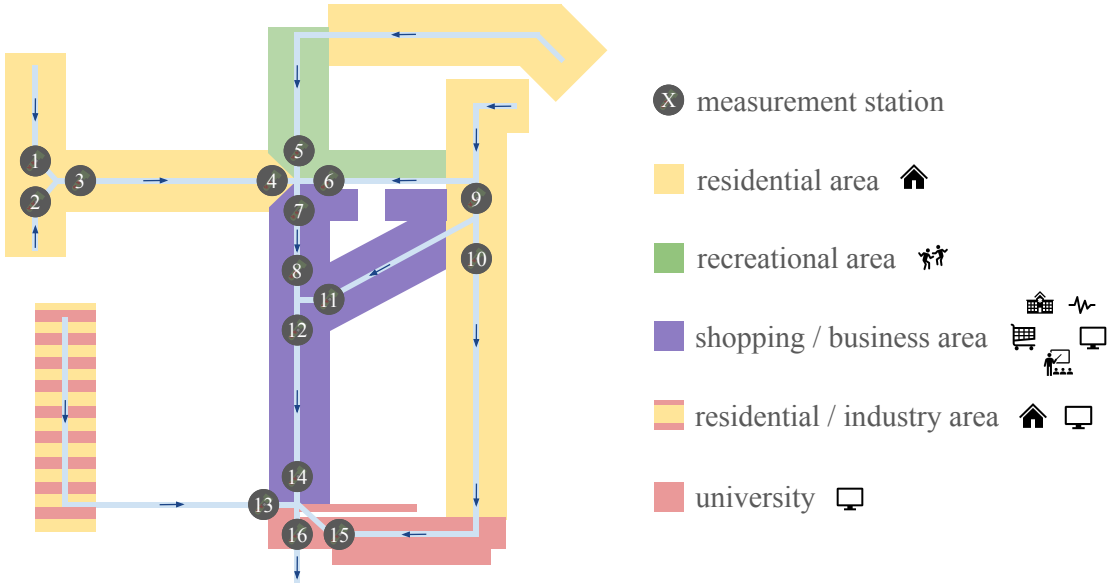


Figure 2: **Demonstrator Neighborhood.** The synthetic neighborhood on which we base our simulation study constitutes areas of different types. Flow rates and substance concentrations are simulated for 16 different measurement stations.

simulated movements, both on weekdays and weekends, remain completely inside the model. The model's synthetic sewer system was designed such that realistic sewer conditions are maintained during all simulation scenarios: No sanitary sewer overflow occurs, all pipes are at a maximum of 90% of their hydraulic capacity, and gravity flow is realized throughout the whole system.

Following the official reporting standards for COVID-19 cases in Germany, we considered $n_A = 6$ age groups ranging from small children to seniors. Households – i.e. groups of agents that share their assigned *Home* location – were created for each residential area based on its total number of agents. We considered 1- to 5-person households. Every household had at least one member of the adult age groups $\mathcal{A} \in \{3, 4, 5, 6\}$ (age groups 1 and 2 correspond to early childhood or adolescence). The household distribution was motivated by the German micro census 2019 [40]. Non-*Home* locations also had to be assigned to the agents. A location of a given type was assigned to an agent from an equal distribution of all locations of that type. Every agent was assigned a location of type *Shop*, *Recreation*, *Hospital*, and *ICU*, while a *School* location was only assigned to agents in age group 2 and a *Work* location only to agents in age groups 3 and 4. Finally, the initial infection states were allotted to agents by independent and identical sampling from the initial infection state distribution (0.2% E, 0.5% I_{ns}, 0.29% I_{sy}, and 0.01% I_{sev}), i.e. on average, 1% of the modeled agents were initially infected.

The integrated model contains various parameters describing characteristics of the virus, which allows for the modeling of a broad spectrum of communicable respiratory or, in particular, COVID-19-like diseases. We implemented the simulation using wild-type COVID-19-like parameters based on [25], [22], and [21]. The ++SYSTEMS files defining the exact shape of the sewage system, area characteristics, etc. are provided as supplementary material for each experiment and in the subsequent sections we only mention settings that differ between the experiments. An overview of the ABM parameters as well as the experiment-specific parameters is provided in Appendix C.

Since most cluster systems are based on Linux, we facilitate the modules using Ubuntu. ++SYSTEMS is a Windows program, hence, we created a headless virtual machine, which can be started and navigated through via a command line interface. One ABM simulation takes about 3.8 minutes on one core; one ++SYSTEMS simulation takes about 3-4 minutes on 8 cores.

3.2. Non-trivial relation of prevalence in catchment area and measurement concentrations

For a first assessment of the process dynamics, we considered the total number of upstream agents, the number of upstream infected agents, and the upstream RNA influx for a single measurement station for an outbreak scenario (Fig. 3). The model simulation reveals that the number of agents in a particular catchment area changes over the course of a day and from weekdays to weekends, primarily due to the agents' participation at work, school, or recreational events. This mobility results in substantial changes to the upstream RNA influx. The measured virus levels are additionally influenced by the sewage volume, which itself depends on the total number of agents in the catchment area. Since agents return home from school, etc. at slightly different time points, the virus levels can show large deviations from an average value for only a few simulation minutes. Overall, the model highlights the impact of mobility on wastewater-based surveillance results.

3.3. Characteristics of catchment area influence dynamics

Establishing a monitoring system for wastewater is time- and energy-consuming. Legal permits for measurements and access to the locations have to be organized, the sampling stations have to be set up, and the samples have to be collected and transported to laboratories for further analysis on a regular basis. Hence, it is not surprising that even established monitoring networks have limited sampling locations and time schedules, often reporting 1-3 values per week and neighborhood or city. This renders the optimal placement of sampling locations and the selection of appropriate sampling strategies critical. Here, we investigated the impact of the choice of sampling location using our fine-grained integrated model, as a corresponding real-world study would be infeasible.

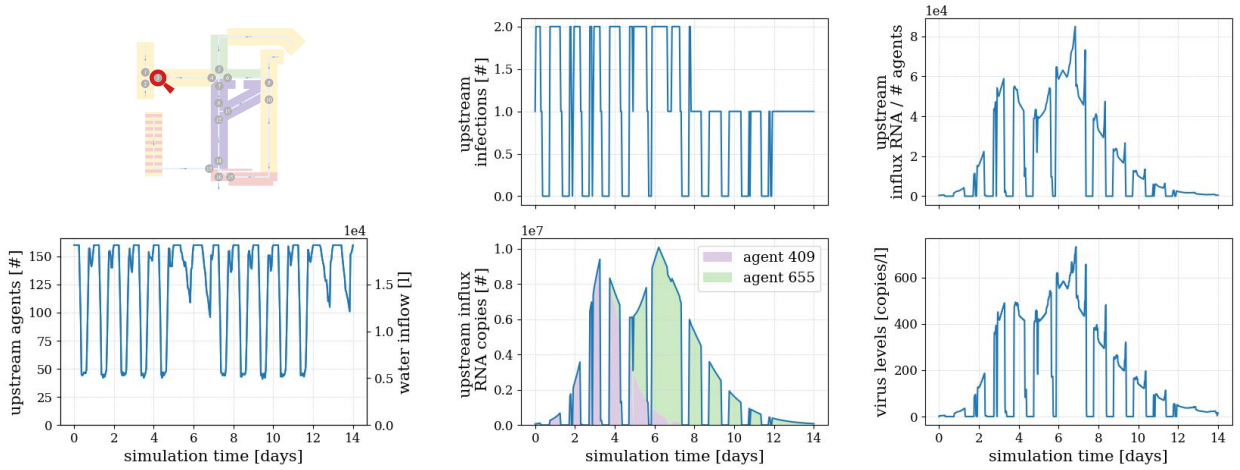


Figure 3: **Exemplary (processed) Simulation Output.** Output for viral measurements in a scenario without precipitation or viral decay at measurement station 3.

We considered 16 possible sampling location in the synthetic neighborhood. These sampling locations correspond to catchment areas with a broad spectrum of different properties: First, the corresponding catchment areas of a sampling station differ with respect to the area type, i.e. primarily include residential areas (stations 1, 2, 3, 4, 9, 10), recreational areas (station 6), shopping/business areas (station 11), or mixed areas. Second, the catchment areas differ with respect to their sizes. Further upstream locations (e.g. station 1) summarize the dynamics of a smaller area than downstream locations (e.g. station 16). In this neighborhood, sewer flow times to the furthest downstream stations are at most around 80 minutes.

To assess the information content of the wastewater-based surveillance data, we conducted a comprehensive simulation study. A total of 250 simulations of the proposed integrated model were used to account for the inherent stochasticity of infection processes; see Supplementary Fig. B.4 for a visualization of the prevalence over time. Assessment of the simulation results (Fig. 4) shows that sampling locations downstream of residential areas (e.g. station 1) produce reproducible daily and weekly trends in the measured virus levels. In contrast, sampling locations downstream of regions containing recreational areas (e.g. station 6 & 11) show more variability between simulations. Sampling stations near the endpoint of the network, which have large catchment areas (e.g. station 16) and would in practice fall closer to a wastewater treatment plant, yield smoother curves with no or less extreme daily and weekly trends.

To evaluate how representative the viral load in the wastewater is at the different sampling locations, we computed the temporal cross-correlations between the RNA copies per liter in wastewater samples and (i) the true overall prevalence (Fig. 5(a), top) and (ii) the true viral shedding into the wastewater (Fig. 5(a), bottom). We found the highest cross-correlation

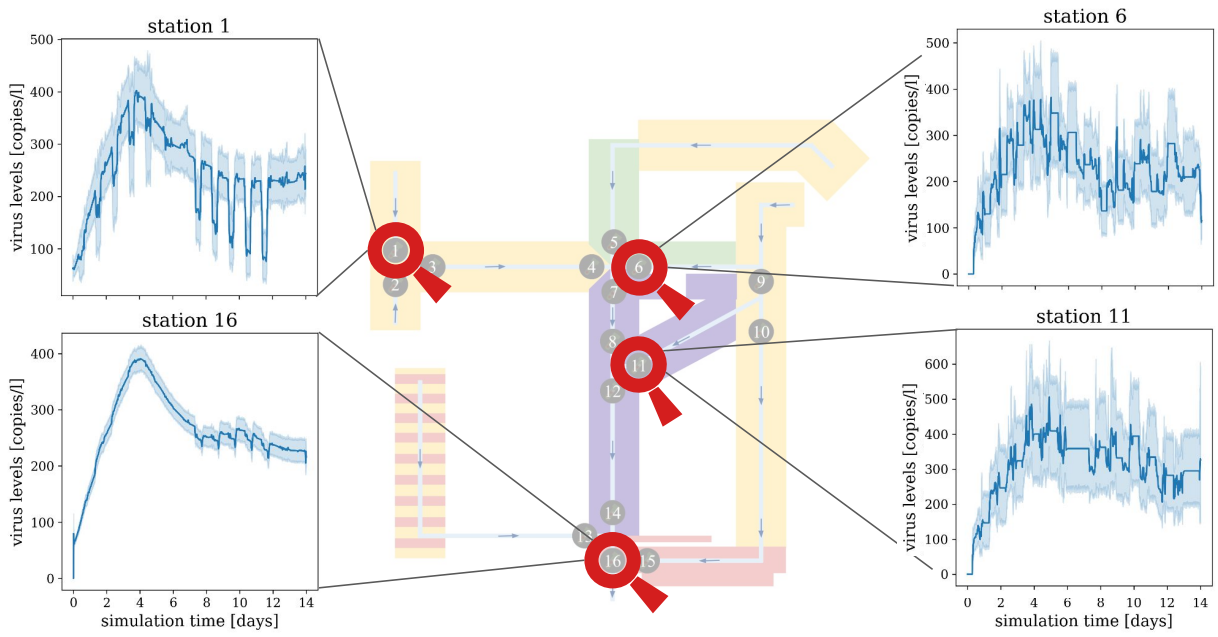


Figure 4: **Variability of Measurements for Sampling Locations.** Comparison of virus levels in wastewater at different sampling locations. Shown are the mean (solid line) and 95% confidence intervals (shaded area) of the 250 simulation results per location.

values are reached for the stations with larger catchment areas, in particular stations 7, 8, 12, 14, and 16. The correlation coefficient is as high as 0.56 for the true prevalence and 0.90 for the amount of virus shed. These corelation coefficients are surprisingly high given the variability between the (stochastic) simulation runs (Fig. 5(b)). Indeed, if the initial infections were not distributed randomly, the pattern would be more pronounced and large catchment areas would be even more beneficial (results not shown).

The integrated model also shows that the temporal cross-correlation is generally higher when a negative time lag is applied to the (true) prevalence data. Since a high virus concentration in the wastewater indicates that the level of infectiousness across the population is also high and that a wave of new infections will therefore likely soon follow, the virus level at sampling stations with large catchment areas was most predictive for the prevalence 10 to 40 hours later. This time lag likely depends on the incubation time of the virus and its replication rate in the human body. As the reported prevalence is delayed compared to the true one, the time shift observed in practice will be even larger (Fig. 5(c)). Overall, our results suggest that in the absence of complicating factors such as viral decay – the impact of which would be limited in this particular sewer due to the relatively short travel times – choosing a wastewater sampling location far enough downstream to be unaffected by daily and weekly trends may help predict increases in prevalence before they occur.

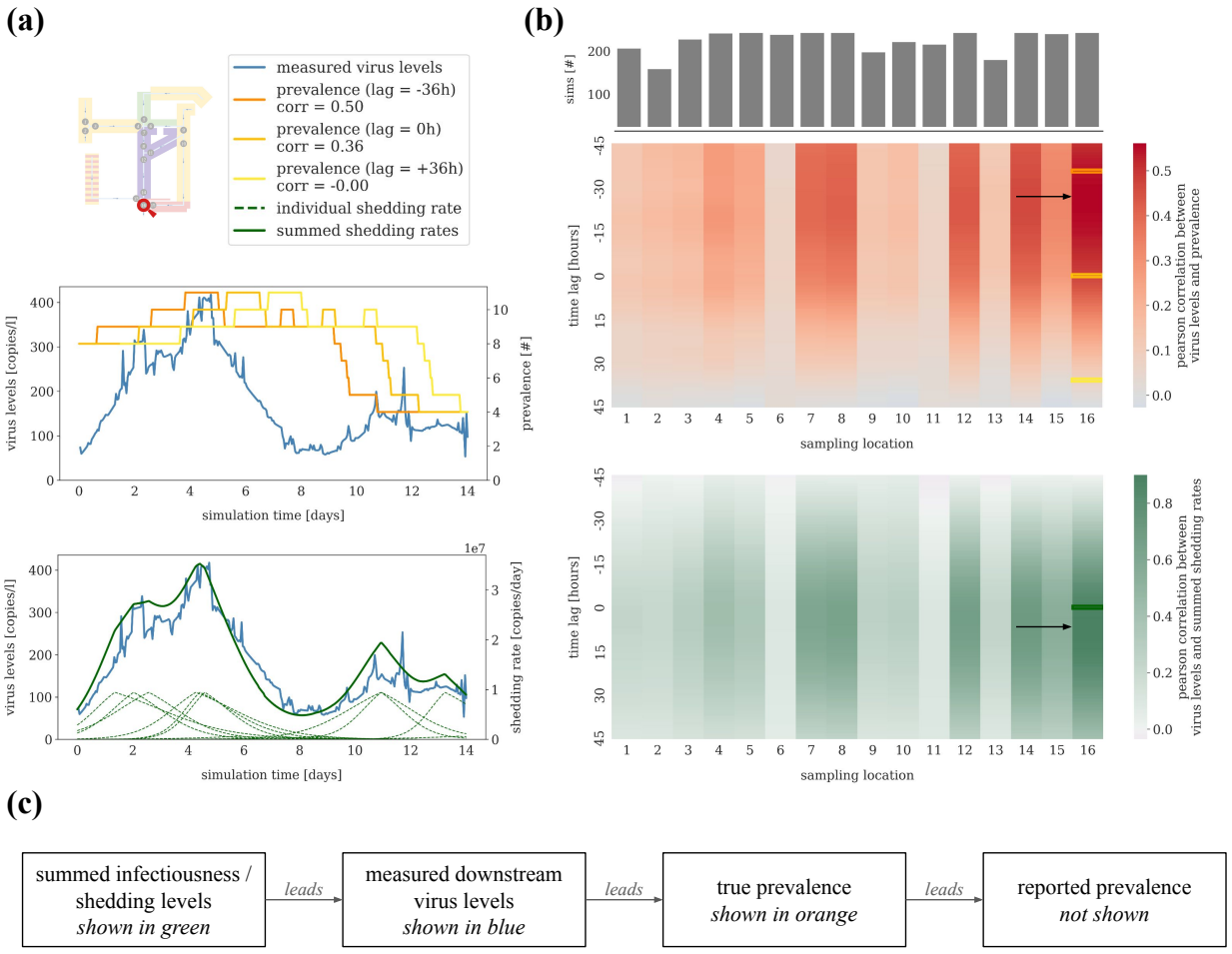


Figure 5: **Cross-Correlations Between Wastewater-based Surveillance Data and Prevalence.** (a) Trajectory of the wastewater viral load in RNA copies per liter for sampling location 16 compared to the total true prevalence shifted with lags -36, 0, and 36 hours (top), or to the summed shedding rates across all prevalent infections (bottom), for one simulation. (b) Pearson cross-correlations between RNA copies per liter in wastewater measured at the 16 different locations and the total true prevalence (top) or the summed shedding rates (bottom), averaged over 250 simulations. The time lag describes the shift in prevalence or shedding rates. The maximum cross-correlations are marked with black arrows. The upper bar plot indicates how many simulations were used to calculate the correlations for each sampling location; simulations in which no virus was ever measured at a particular location were removed. (c) Schematic illustrating the temporal relationship between the different outcomes.

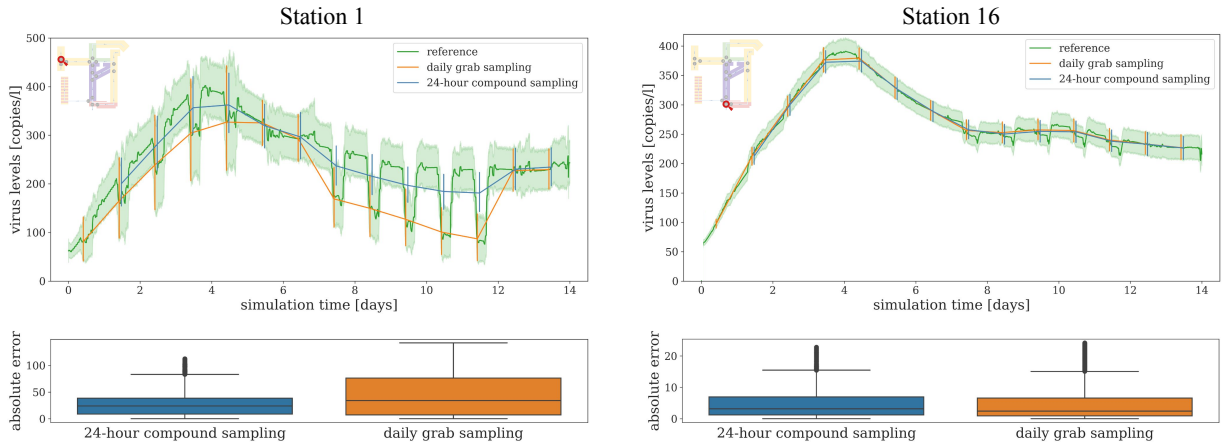


Figure 6: **Sampling Protocols.** RNA concentration in wastewater with 24-hour compound sampling or daily grab sampling (at 10:00 am each day) compared to the reference scenario (grab sampling every three minutes). Shown are the mean (solid lines) and 95% confidence intervals (shaded areas or error bars) of the 250 simulation results for stations 1 and 16 along with the distribution of the absolute error between the linearly interpolated average results for 24-hour compound or daily grab sampling and the average results for the reference scenario.

3.4. Temporal sampling design has minimal impact on wastewater monitoring results if samples are taken downstream

The sampling design differs between wastewater monitoring studies. The most common setups are one grab sample collected per day of interest (usually during the morning flush) or a 24-hour compound sample based on a collection of one sample per hour [23]. To assess the impact of these sampling strategies and their benefits and disadvantages, we simulated both strategies using the same setup as in the previous section (in which we assumed the use of discrete grab samples every three minutes). The analysis of the simulation results indicates that in the case of a clear daily trend, the choice of sampling protocol can influence the results and e.g. lead to systematically biased estimates of the general dynamics (Fig. 6). Stations further downstream are less influenced by daily and weekly agent movement patterns (see Section 3.3) and different sampling protocols yield comparable results, i.e. a maximal cross-correlation of 0.56 between measured viral load and time-lagged prevalence.

3.5. Rain influx impacts reliability of wastewater monitoring results in a nonlinear manner

Rain influences the amount of fluid in the wastewater system, the fluid velocity, and the concentration of particles in the overall wastewater. Yet, many currently available models simply disregard rain events and the associated dataset. This approach results in a loss of information, and – as it is unclear how long-lasting the effects of rain might be – might still not be particularly reliable. To provide a fine-grained analysis of the impact of rain

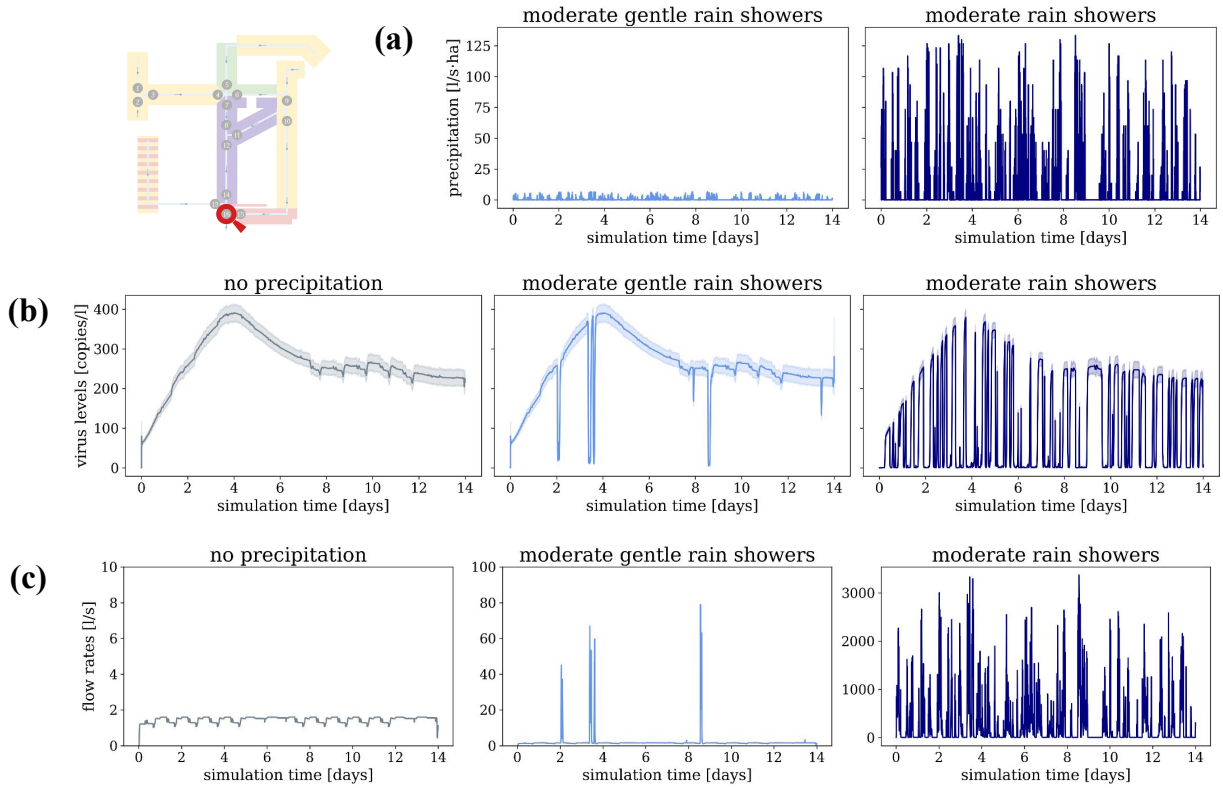


Figure 7: **Influence of Precipitation.** Comparison of different rain scenarios (no precipitation, moderate gentle rain showers, and moderate rain showers) at sampling location 16 visualizing (a) the precipitation, (b) the RNA concentration in wastewater in copies per liter, and (c) the flow rates in liter per second.

on wastewater measurements, we simulated three scenarios: *No precipitation* (which was also used for the previous results), *moderate gentle rain*, and *moderate rain* showers. Here, we followed the rain intensity definitions from Germany's National Meteorological Service (DWD) [42]: "moderate gentle rain" means between 0.1 mm and 0.5 mm in 60 min and "moderate rain" means between 2.5 mm and 10.0 mm in 60 min.

In order to run simulations with a two-week duration with realistic time-dependent rain intensities, a suitable two-week period from a synthetic rain series [3] generated by the Bavarian Environment Agency (LfU) was used. The intensities were adapted such that the above-mentioned rain definition criteria were met. As a result, the two rain scenarios only differ in their intensities; the temporal profile of rain peaks is the same for both scenarios, ensuring comparability.

Due to effects like evaporation, the filling of water basins (e.g. uptake by the ground), and permeable and non-permeable surface fractions, only a small proportion of rainfall ends up in the hydraulic system and a minimum amount of precipitation is necessary to have an effect at all. This net hydraulic surface runoff increases the water volume and hence, can increase the

flow rates and reduce the concentration of RNA in wastewater significantly (Fig. 7). RNA copies per liter close to zero correspond to a proportion close to 1 of rainwater in the sewage. If the flow rates increase by several orders of magnitude for a short amount of time (as is, e.g. the case in the moderate gentle rain scenario presented here), the sewage containing large fractions of rainwater is flushed out of the system very quickly, yielding a state comparable to the non-precipitation state afterwards. The moderate gentle rain scenario showcases that there is a minimum amount of precipitation necessary to have an effect on the flow rates and virus concentrations. The moderate rain scenario highlights how precipitation influences measurements if the rainwater inflow to the sewage is larger than zero for several neighboring time steps.

The analysis of the influence of precipitation events on measurements showcases the importance of normalizing observations to compare measurements of precipitation and dry weather time points. For more details on normalization strategies, see the results presented in Section 3.7.

3.6. Virus Characteristics

The interplay of virus and host immune response determine virus shedding and transmission and, hence, the prevalence and influx of virus particles into the wastewater system. Yet, the virus particles are not necessarily stable but can decay. For SARS-CoV-2, estimates of the the 90% reduction times in wastewater at ambient temperatures range between 5.5 and 28.8 days [5, 13, 1]. As the sewer of the synthetic neighborhood has a relatively short maximum flow time, we have been able to reasonably neglect viral decay so far. We now assume that our virus of interest has much faster virus reduction times than SARS-CoV-2, in order to study the impact of rapid decay processes on wastewater monitoring results and their relation to prevalence. To this end, we compare three different temporal models for the decay rate v of [RNA] (in copies per liter) in wastewater:

- no decay: $v = 0 \ \frac{\text{copies}}{l \cdot s}$
- linear decay: $v = k_1, \ \text{where } k_1 = \begin{cases} -0.1 \frac{\text{copies}}{l \cdot s}, \ \text{for [RNA]} > 0 \\ 0 \frac{\text{copies}}{l \cdot s}, \ \text{for [RNA]} \leq 0 \end{cases}$
- exponential decay: $v = k_2 \cdot [\text{RNA}], \ \text{where } k_2 = -1 \cdot 10^{-3} \frac{1}{s}$ (corresponding to a half-life of about 690 seconds or about 0.2 hours)

The *no decay* scenario assumes that viral particles remain intact and serves as a baseline for understanding the upper bound and comparing it to more realistic models. The *linear decay* scenario is motivated by a potential interaction between a virus and a certain enzyme or

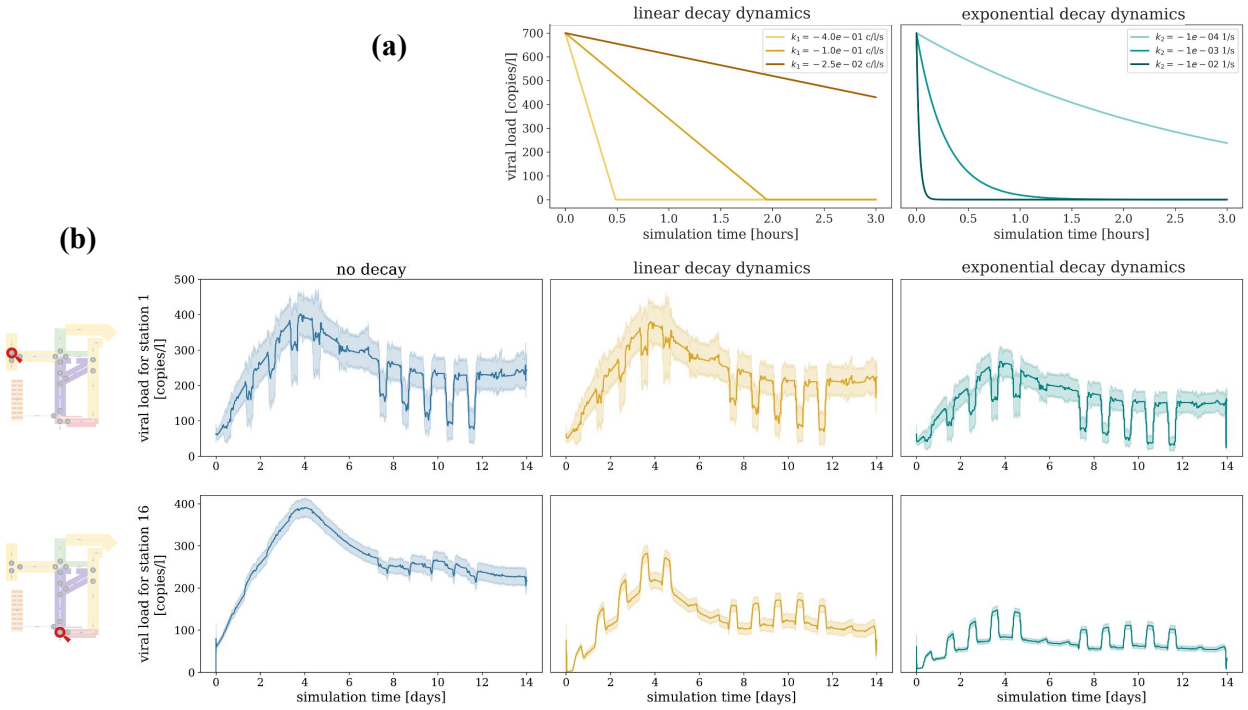


Figure 8: **Viral Decay Dynamics.** (a) Illustration of the effects of linear decay (left) vs. exponential decay (right) with 3 example parameter settings each on a starting virus concentration of 700 copies per liter. (b) The virus concentration measured at station 1 (top) and station 16 (bottom) over time without rain and with no decay (left), linear decay with $k_1 = -0.1 \frac{\text{copies}}{l \cdot s}$ (center), and exponential decay with $k_2 = -1 \cdot 10^{-3} \frac{1}{s}$ (right).

environmental condition in the wastewater that degrades the virus at constant rate (e.g. because its abundance is limited). The *exponential decay* scenario captures the most commonly observed decay dynamics, translating to a constant decay probability per unit time.

We analyzed the simulation results for the three decay scenarios, without precipitation, at two sampling locations: 1 and 16 (Fig. 8). At the upstream station 1, the measured viral loads are proportionally slightly lower for the linear decay setting and considerably lower for the exponential decay setting compared to the no decay setting, since for the relevant concentrations, exponential decay with a half-life of 690 seconds is considerably faster than linear decay of -0.1 copies per liter per second. However, the general shape of the viral load trajectory at station 1 over time is unaffected by the decay setting: For all three decay scenarios, there are periodic dips in the virus level on weekdays and a defined peak around the fourth day. At the downstream station 16, not only are the measured viral loads lower for the linear and especially the exponential decay settings, but the shape of the viral load trajectory over time is also affected, with defined peaks during weekday daytime periods. When the viral decay is non-negligible, virus copies shed from the upstream residential areas tend to decay

before they reach the furthest downstream station, so station 16 primarily measures the copies shed from the nearest areas: a university and a shopping/business region, which are only inhabited during working hours. This phenomenon indicates that the interaction between viral load and viral decay can be complex and that a detailed sewer model is necessary to adequately describe its impact on the reliability of wastewater viral loads as an indicator of disease prevalence.

3.7. Virus Normalization

Wastewater systems can have – as outlined above – variable flow rates depending on factors like precipitation and water usage. Hence, wastewater samples have to be normalized to ensure accurate and reliable data interpretation. The two most commonly used normalization strategies are based on either flow rates or additional indicators like the concentration of Pepper Mild Mottle Virus (PMMoV). PMMoV is a plant virus commonly found in human feces at relatively stable concentrations and hence, serves as a good indicator for the amount of human waste in the sample. Here, we use the model to assess which normalization strategy yields corrected viral load values closest to the those one would measure if there was no precipitation event.

The normalization is calculated using one of the following two strategies:

- normalization with flow rates: $[\mathrm{RNA}]_{normalized} = \frac{Q}{\mathrm{mean}(Q_{dry})} \cdot [\mathrm{RNA}]$
- normalization with PMMoV: $[\mathrm{RNA}]_{normalized} = \frac{\mathrm{mean}([\mathrm{PMMoV}]_{dry})}{[\mathrm{PMMoV}]} \cdot [\mathrm{RNA}]$

where Q is the current flow rate, $\mathrm{mean}(Q_{dry})$ the mean flow rate on dry days, [PMMoV] the current PMMoV concentration, and $\mathrm{mean}([\mathrm{PMMoV}]_{dry})$ the mean concentration of PMMoV on dry days [31]. Rainwater infiltration into a sewer system dilutes PMMoV while increasing flow rates, so correcting wastewater measurements using the ratio between the expected and measured PMMoV concentration or between the measured and expected flow rate can help reduce unwanted variability in wastewater-based data.

We followed exactly these procedures using simulated data. To simulate measurements of PMMoV we assumed a constant PMMoV shedding per agent throughout the simulations and that PMMoV is never subject to viral decay. A small number of observations (between 0.15% and 0.28% per scenario), for which the measured PMMoV concentration was zero copies per liter, were removed. As seen in Fig. 9(b), flow rate normalization (right column) is not effective for correcting wastewater measurements in our model for the effects of either moderate gentle or moderate rain. In both the moderate gentle and moderate rain scenarios, infiltration of rainwater into the model sewer system causes greater proportional increases to the flow rates than decreases to the virus concentrations, so normalization with flow rates

leads to notable over-corrections. In contrast, normalization with PMMoV (center column) appears highly effective. Rainwater infiltration affects PMMoV and the virus of interest similarly, so the PMMoV-normalized measurements, while noisier, generally match the no-rain reference scenario in terms of both shape and scale. Specifically, applying PMMoV normalization reduces the median (across all simulations and time points) of the absolute error between the virus concentration at station 16 for the moderate rain, no decay scenario and the no rain, no decay reference scenario from about 113 to about 8 copies per liter (see Fig. 9(d) and Supplementary Table C.5). Since our PMMoV normalization approach corrects only for rainfall and not for viral decay, PMMoV normalization only marginally improves the median absolute error compared to the reference scenario for the moderate rain, exponential decay and the moderate rain, linear decay scenarios. However, PMMoV appears even less effective when the viral decay is linear rather than exponential. This is likely because in the linear decay scenario, unlike the exponential decay scenario, viral concentrations can and do degrade to zero, which renders normalization useless for certain observations.

Rain reduces the cross-correlations between overall prevalence in the catchment area and the measured RNA copies per liter in wastewater samples over time (see Fig. 10). For the scenarios with rain but without viral decay, applying PMMoV normalization to the wastewater measurements restores the cross-correlation coefficients to very close to their levels in the reference scenario. For the scenarios with both rain and viral decay, PMMoV normalization partially restores the general trends in viral load measurements over time, and therefore has a more limited but non-negligible impact on cross-correlations.

4. Discussion

In this study, we presented a first integrative model for the fine-grained description of infectious disease dynamics and wastewater surveillance. The model couples a stochastic model of individual mobility, infection transmission, and disease progression with a highly detailed hydrodynamic model of viral RNA transport through wastewater networks. Using this model for the description of a synthetic neighborhood and corresponding sewer system, we were able to investigate the influence of sampling protocols, precipitation events, virus characteristics, and normalization strategies on the relationship between infection dynamics and resulting wastewater measurements. We found that locations for sampling stations should be chosen carefully, so that they lie downstream of a sufficient number of agents and diverse location types; that precipitation and viral decay can have unexpected, nonlinear impacts on wastewater viral load that require detailed integrative modeling approaches to be understood; and that flow rate normalization should only be implemented with caution as it can lead to large over-corrections if there has recently been precipitation. Overall, our study suggests that if

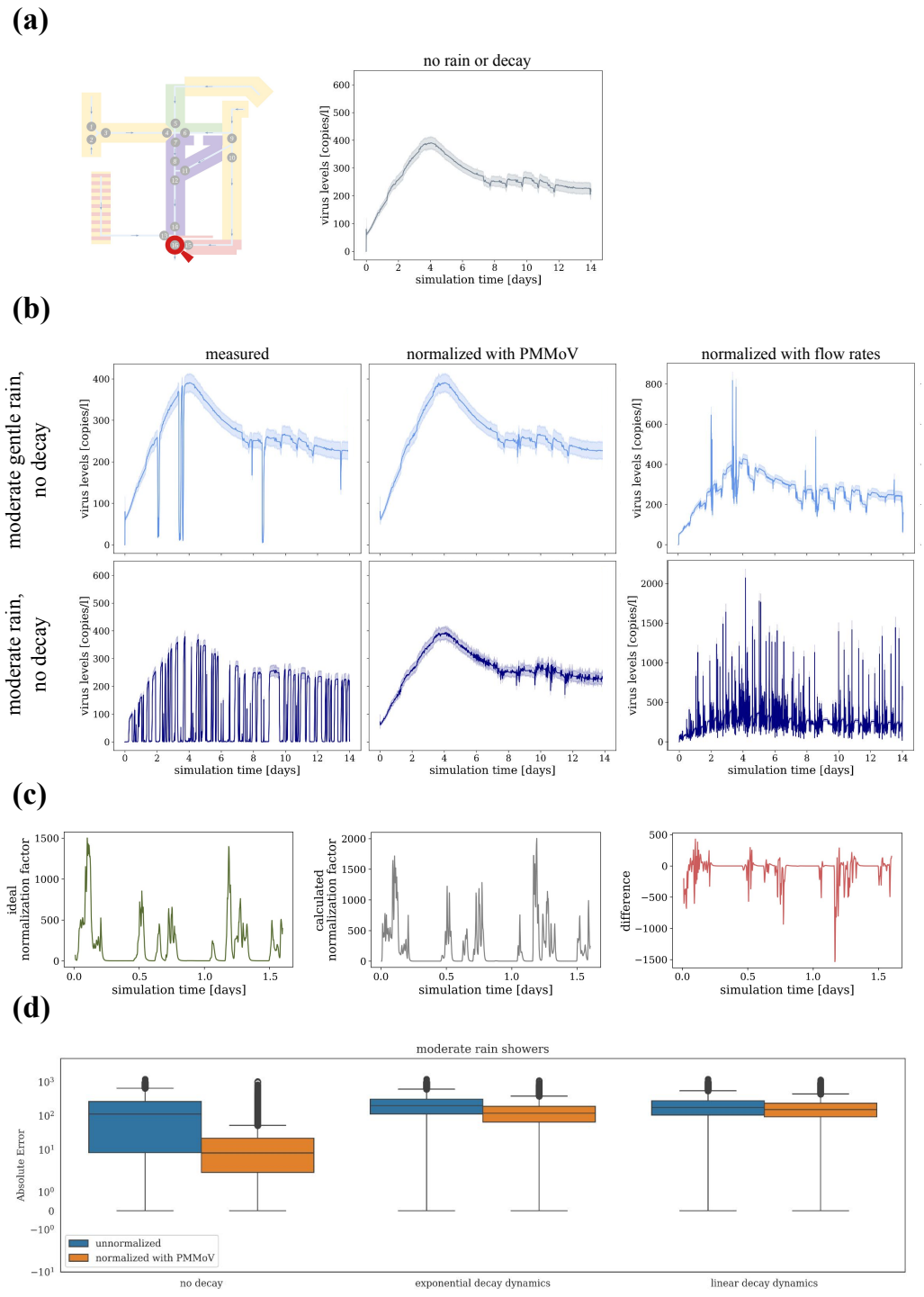


Figure 9: **Normalization Strategies Applied to Simulations.** (a) The virus concentration at station 16 in the no-rain, no-decay reference scenario. (b) The virus concentration at station 16 in the moderate gentle rain (top) or moderate rain (bottom) scenarios, as measured (left), after normalization with PMMoV (center), and after normalization with flow rates (right). (c) Comparison of the "ideal" normalization factor that would transform the moderate rain results to the no-rain reference results, versus the actual flow-rate-based normalization factor, for one example simulation. (d) The distribution over all simulations and time points of the absolute error between the moderate rain results with no, exponential, or linear decay dynamics and the no-rain, no-decay reference scenario.

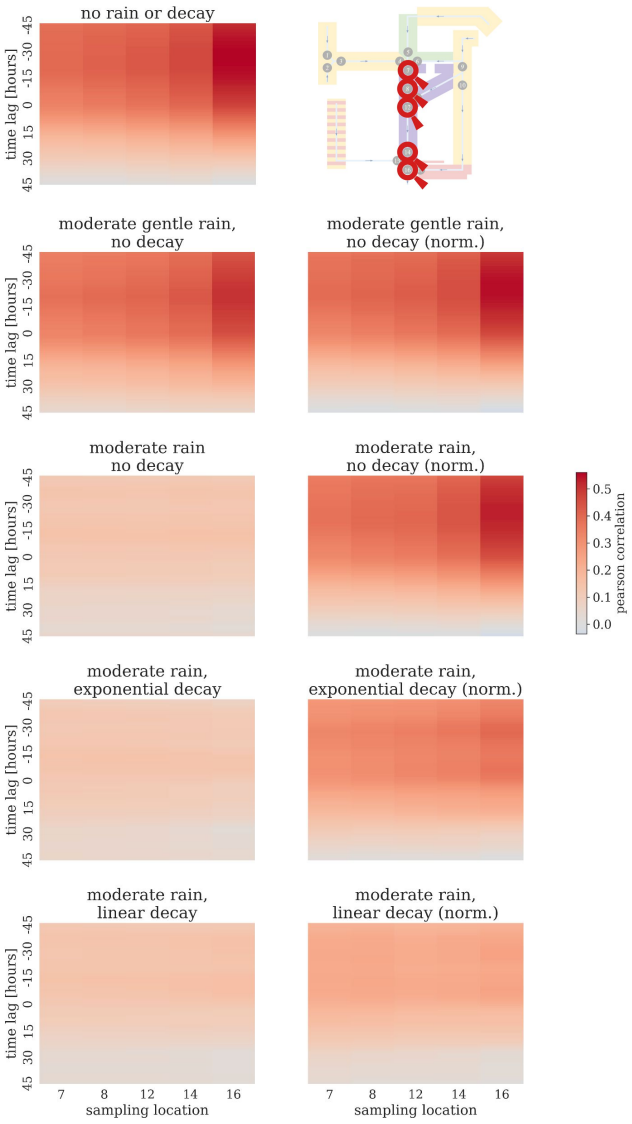


Figure 10: **Cross-Correlations Between Wastewater Samples and Prevalence by Scenario.** Pearson cross-correlations between RNA copies per liter in wastewater and prevalence over time for five downstream stations, averaged over 250 simulations, for the no-rain, no-decay reference scenario (top left), various decay and rain scenarios without normalization (left column), and the same decay and rain scenarios with PMMoV normalization (right column).

appropriate sampling, normalization, and analysis techniques are used, then wastewater may serve as a leading indicator of disease prevalence.

Sampling station characteristics – including the size of the upstream population, the times of day during which the upstream areas are generally populated, and the distribution of sewer travel times to the station – can both qualitatively and quantitatively affect wastewater measurement dynamics. To produce the highest correlations with prevalence, our study indicates that sampling stations should be placed far enough downstream to receive wastewater from a representative sample of the population of interest as well as from a mix of homes, places of work and study, and recreational areas. If infections are distributed throughout the catchment area, then 24-hour compound sampling can help alleviate some of the disadvantages of upstream sampling location placement, but if infections are localized, downstream placement becomes more crucial, as a sampling station too far upstream may miss the outbreak entirely. However, these considerations must be balanced against others: our study showed how the effects of viral decay become more pronounced the further downstream a sampling station is, and in real-world scenarios, inflow of industrial wastewater may also be a concern. Thus, while we generally argue for mid- or downstream sampling locations, a detailed modeling approach like ours is needed to choose the optimal sampling location for a particular region of interest.

Our study also illustrates how precipitation and viral decay, separately or together, can have complex and nonlinear impacts on the measured viral load in wastewater, which need to be accounted for when using wastewater data as a public health indicator. Due to the interactions between evaporation, water retention, and other factors, the influence of rainfall on sewer flow rates and viral concentrations within different pipes can be difficult to predict; for example, our model shows how a continuous drizzle can potentially lead to discrete drops in wastewater measurements. Although using sensors to measure sewer flow rates at a sampling station and then adjusting wastewater measurement accordingly is possible, this approach performed sub-optimally in our study and often led to dramatic over-corrections. We find that normalization with a human fecal indicator such as PMMoV may be preferable in sewer systems that have recently been infiltrated by rainwater. This finding is in contrast to the conclusions of Rainey et al. [36], who recommended normalization with flow rates to account for variations in the size of a sewershed's service population. As Rainey et al. did not consider the effects of precipitation, future work should explore how best to account for both precipitation and population size when normalizing wastewater measurements for comparisons both over time and across sampling locations.

Like precipitation, viral decay in our study led to both qualitative and quantitative changes in the viral load trajectory measured at a particular station over time. Our decay scenarios were intentionally exaggerated – we considered exponential decay with a 90% reduction time of about 0.6 hours, whereas the actual 90% reduction time of SARS-CoV-2 in wastewater is

temperature-dependent but usually estimated to be upwards of five days – but this extreme case illustrates how the signal from surface areas far from the sampling location can be lost if viral decay is not appropriately accounted for. In our study, we found that even in the presence of rapid viral decay, PMMoV normalization could restore non-negligible correlations between prevalence and wastewater measurements, but it could only slightly reduce the absolute error relative to a no-rain, no-decay reference scenario. Since PMMoV is very stable in wastewater, normalization with PMMoV alone cannot account for the effects of viral decay on wastewater measurements. Instead, an appropriate viral decay model should be chosen based on the characteristics of the virus of interest and the sewer system, including temperature [14] and biofilms [44] – both of which are likely to be affected in turn by the amount of precipitation in the pipes. Due to its level of detail, our study provides new insights into the interactions between rain, viral decay, and wastewater measurements and underscores the importance of appropriate normalization and analysis of wastewater data.

Overall, our study indicates that despite potential confounding factors, if appropriate sampling, normalization, and analysis techniques are utilized, then wastewater-based surveillance data can provide insights into trends in disease prevalence and possibly predict outbreak peaks 1 to 2 days before they occur. Viral load measurements in wastewater depend not only on the total number of people shedding, but also on the viral load – and, by extension, the infectiousness – of each prevalent infection. High viral loads in wastewater indicate high infection potential across the catchment area, meaning that new infections are likely to occur soon; thus, in our model, the peak in viral load measurements at the furthest downstream station tended to occur about 30 hours before the corresponding peak in overall prevalence. Previous studies, such as the one by Peccia et al. [34], have found that epidemiological measures such as positive test counts and hospital admissions tend to lag several days behind wastewater measurements, and our study suggests that this may not be entirely due to reporting delays. Thus, our results support Peccia et al.'s conclusion that wastewater-based surveillance data can help guide public health officials in deciding when to implement or ease infection control measures.

4.1. Limitations and Future Work

One limitation of our model is that we assumed each agent's water usage to be distributed uniformly across the day. In reality, an individual would only produce wastewater – and, if infected, shed into the sewer system – at certain time points, e.g. while using the bathroom at home after waking up in the morning. We expect that realistic patterns of water usage over time would likely create additional daily trends in wastewater measurements and highlight the advantages of 24-hour compound sampling over grab samples, but future work on this model should incorporate the circadian rhythm of water consumption and shedding behavior to confirm this.

We also leave for future work a sensitivity analysis of how relaxing the assumptions underlying our shedding model might affect our results. Reinfections, vaccinations, and cross-talk between co-circulating pathogens were out of the scope of this initial model, and we therefore did not consider, for example, how the effects of vaccinations on viral shedding [11] could complicate the relationship between disease dynamics and wastewater measurements. Due to data limitations, we assumed that viral shedding in urine and feces was proportional to respiratory shedding and that the peak viral load value was the same for all symptomatic infections, although realistic variations in this value across individuals might decrease the cross-correlations between prevalence and wastewater measurements when the number of infections is small.

Finally, the methods we have so far only applied in the context of an synthetic neighborhood should, in the future, be adapted to real-world scenarios. This will require, for example, introducing an appropriate noise model to account for the effects of wastewater measurement uncertainty and detection limits. Thus, our model's potential ability to map real-world wastewater measurements back to underlying prevalence remains to be tested.

4.2. Conclusions

Our study illustrates the value of sophisticated models of infection and wastewater dynamics and highlights the potential of wastewater-based surveillance data to reflect trends in prevalence without being influenced by sampling bias, reporting delays, or under-ascertainment. While applications to real-world data remain for future work, our simulation study provided key insights into the advantages of downstream sampling location placement, 24-hour compound sampling, models for the effects of viral decay, and PMMoV normalization.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC 2047 – 390685813, EXC 2151 – 390873048), by the German Federal Ministry of Education and Research (BMBF) (INSIDe – grant numbers 031L0297A, 031L0297B, 031L0297D, and 031L0297E), and by the University of Bonn (via the Schlegel Professorship of J.H.). It was furthermore supported by the Initiative and Networking Fund of the Helmholtz Association (grant agreement number KA1-Co-08, Project LOKI-Pandemics) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (grant agreement 528702961).

Author Contributions

N.S. – conceptualization, methodology, formal analysis, investigation, writing — original draft, writing — review & editing, visualization; J.B. – methodology, software, writing — original draft, writing — review & editing; A.F.H. – methodology, software, writing — original draft, writing — review & editing, funding acquisition; K.W.S – formal analysis, writing — original draft, writing — review & editing, visualization; D.K. – methodology, software, writing — review & editing; A.W. – validation, writing — review & editing, supervision, funding acquisition; M.J.K – validation, writing — review & editing, supervision, funding acquisition; J.H – validation, writing — review & editing, supervision, project administration, funding acquisition.

Implementation and Availability

The version of MEmilio used in this study, including all input files, is publicly available at `https://github.com/SciCompMod/memilio/tree/inside-demonstrator-final`. Access to the software ++SYSTEMS and its documentation is available upon reasonable request to andreas.hofmann@tandler.com; all parameter files and simulation outputs used in this study are available at `https://doi.org/10.5281/zenodo.14046493`. Results were analyzed using the Python code available at `https://github.com/inside-consortium/inside_demonstrator`.

References

[1] Warish Ahmed, Paul M. Bertsch, Kyle Bibby, Eiji Haramoto, Joanne Hewitt, Flavia Huygens, Pradip Gyawali, Asja Korajkic, Shane Riddell, Samendra P. Sherchan, Stuart L. Simpson, Kwanrawee Sirikanchana, Erin M. Symonds, Rory Verhagen, Seshadri S. Vasan, Masaaki Kitajima, and Aaron Bivins. Decay of SARS-CoV-2 and surrogate murine hepatitis virus RNA in untreated wastewater to inform application in wastewater-based epidemiology. *Environmental Research*, 191, 2020.

[2] Peter M. Bach, Wolfgang Rauch, Peter S. Mikkelsen, David T. McCarthy, and Ana Deletic. A critical review of integrated urban water modelling – urban drainage and beyond. *Environmental Modelling & Software*, 54:88–107, 2014.

[3] LfU Bayern. NiedSimBy: Synthetische Niederschlagsreihen - LfU Bayern, 2024. `https://www.lfu.bayern.de/wasser/abwasser_synthetische_niederschlagsreihen/index.htm` (visited on 02/20/2024).

[4] Julia Bicker, René Schmieding, Michael Meyer-Hermann, and Martin J. Kühn. Hybrid metapopulation agent-based epidemiological models for efficient insight on the individual scale: a contribution to green computing. 2024. Submitted for publication. `https://arxiv.org/abs/2406.04386`.

[5] Jean-Baptiste Burnet, Henry-Michel Cauchie, Cécile Walczak, Nathalie Goeders, and Leslie Ogorzaly. Persistence of endogenous rna biomarkers of sars-cov-2 and pmmov in raw wastewater: Impact of temperature and implications for wastewater-based epidemiology. *Science of The Total Environment*, 857, 2023.

[6] François M. Castonguay, Julie C. Blackwood, Emily Howerton, Katriona Shea, Charles Sims, and James N. Sanchirico. Optimal spatial evaluation of a pro rata vaccine distribution rule for covid-19. *Scientific Reports*, 13(1):2194, 2023.

[7] Min-Kyung Chae, Dong-Uk Hwang, Kyeongah Nah, and Woo-Sik Son. Evaluation of COVID-19 intervention policies in South Korea using the stochastic individual-based model. *Scientific Reports*, 13(1):18945, 2023.

[8] Daniel Champredon, Irena Papst, and Warsame Yusuf. ern: An r package to estimate the effective reproduction number using clinical and wastewater surveillance data. *PLoS ONE*, 19(6), 2024.

[9] K.K. Chau, L. Barker, E.P. Budgell, K.D. Vihta, N. Sims, B. Kasprzyk-Hordern, E. Harriss, D.W. Crook, D.S. Read, A.S. Walker, and N. Stoesser. Systematic review of wastewater surveillance of antimicrobial resistance in human populations. *Environment International*, 162:107171, 2022.

[10] Lorenzo Contento, Noemi Castelletti, Elba Raimúndez, Ronan Le Gleut, Yannik Schälte, Paul Stapor, Ludwig Christian Hinske, Michael Hoelscher, Andreas Wieser, Katja Radon, Christiane Fuchs, and Jan Hasenauer. Integrative modelling of reported case numbers and seroprevalence reveals time-dependent test efficiency and infectious contacts. *Epidemics*, 43:100681, 2023.

[11] Miguel Garcia-Knight, Khamal Anglin, Michel Tassetto, Scott Lu, Amethyst Zhang, Sarah A Goldberg, Adam Catching, Michelle C Davidson, Joshua R Shak, Mariela Romero, Jesus Pineda-Ramirez, Ruth Diaz-Sanchez, Paulina Rugart, Kevin Donohue, Jonathan Massachi, Hannah M Sans, Manuella Djomaleu, Sujata Mathur, Venice Servellita, David McIlwain, Brice Gaudiliere, Jessica Chen, Enrique O Martinez, Jacqueline M Tavs, Grace Bronstone, Jacob Weiss, John T Watson, Melissa Briggs-Hagen, Glen R Abedi, George W Rutherford, Steven G Deeks, Charles Chiu, Sharon Saydah, Michael J Peluso, Claire M Midgley, Jeffrey N Martin, Raul Andino, and J Daniel Kelly. Infectious viral shedding of SARS-CoV-2 Delta following vaccination: A longitudinal cohort study. *PLoS Pathogens*, 18(9):e1010802, September 2022.

[12] Esayas Kebede Gudina, Kira Elsbernd, Daniel Yilma, Rebecca Kisch, Karina Wallrafen-Sam, Gemeda Abebe, Zeleke Mekonnen, Melkamu Berhane, Mulusew Gerbaba, Sultan Suleman, Yoseph Mamo, Raquel Rubio-Acero, Solomon Ali, Ahmed Zeynudin, Simon Merkt, Jan Hasenauer, Temesgen Kabeta Chala, Andreas Wieser, and Arne Kroidl. Tailoring COVID-19 vaccination strategies in high-seroprevalence settings: Insights from Ethiopia. *Vaccines*, 12(7):745, 2024. PubMed-not-MEDLINE.

[13] Ying Guo, Yanchen Liu, Shuhong Gao, Xu Zhou, Muttucumaru Sivakumar, and Guangming Jiang. Effects of temperature and water types on the decay of coronavirus: A review. *Water*, 15, 2023.

[14] Olga E. Hart and Rolf U. Halden. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges. *Science of The Total Environment*, 730:138875, August 2020.

[15] Robert Hinch, William J. M. Probert, Anel Nurtay, Michelle Kendall, Chris Wymant, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, Luca Ferretti, Daniel Montero, James Warren, Nicole Mather, Matthew Abueg, Neo Wu, Olivier Legat, Katie Bentley, Thomas Mead, Kelvin Van-Vuuren, Dylan Feldner-Busztin, Tommaso Ristori, Anthony Finkelstein, David G. Bonsall, Lucie Abeler-Dörner, and Christophe Fraser. OpenABM-Covid19—An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. *PLOS Computational Biology*, 17(7):e1009146, July 2021.

[16] B. R. Hodges. Conservative finite-volume forms of the saint-venant equations for hydrology and urban drainage. *Hydrology and Earth System Sciences*, 23(3):1281–1304, 2019.

[17] Laura D. Howe, Kate Tilling, Bruna Galobardes, and Debbie A. Lawlor. Loss to follow-up in cohort studies: Bias in estimates of socioeconomic inequalities. *Epidemiology*, 24(1), 2013.

[18] Elizabeth Hunter, Brian Mac Namee, and John Kelleher. A Hybrid Agent-Based and Equation Based Model for the Spread of Infectious Diseases. *Journal of Artificial Societies and Social Simulation*, 23(4):14, 2020.

[19] Christopher I. Jarvis, Kevin Van Zandvoort, Amy Gimma, Kiesha Prem, Megan Auzenbergs, Kathleen O'Reilly, Graham Medley, Jon C. Emery, Rein M. G. J. Houben, Nicholas Davies, Emily S. Nightingale, Stefan Flasche, Thibaut Jombart, Joel Hellewell, Sam Abbott, James D. Munday, Nikos I. Bosse, Sebastian Funk, Fiona Sun, Akira Endo, Alicia Rosello, Simon R. Procter, Adam J. Kucharski, Timothy W. Russell, Gwen Knight, Hamish Gibbs, Quentin Leclerc, Billy J. Quilty, Charlie Diamond, Yang Liu, Mark Jit, Samuel Clifford, Carl A. B. Pearson, Rosalind M. Eggo, Arminder K. Deol, Petra Klepac, G. James Rubin, W. John Edmunds, and CMMID COVID-19 working group. Quantifying the impact of physical distance measures on the transmission of covid-19 in the uk. *BMC Medicine*, 18(1):124, 2020.

[20] Terry C. Jones, Guido Biele, Barbara Mühlemann, Talitha Veith, Julia Schneider, Jörn Beheim-Schwarzbach, Tobias Bleicker, Julia Tesch, Marie Luisa Schmidt, Leif Erik Sander, Florian Kurth, Peter Menzel, Rolf Schwarzer, Marta Zuchowski, Jörg Hofmann, Andi Krumbholz, Angela Stein, Anke Edelmann, Victor Max Corman, and Christian Drosten. Estimating infectiousness throughout SARS-CoV-2 infection course. *Science*, 373(6551):eabi5273, July 2021.

[21] David Kerkmann, Sascha Korf, Khoa Nguyen, Daniel Abele, Alain Schengen, Carlotta Gerstein, Jens Henrik Göbbert, Achim Basermann, Martin J. Kühn, and Michael Meyer-Hermann. Agent-based modeling for realistic reproduction of human mobility and contact behavior to evaluate test and isolation strategies in epidemic infectious disease spread. 2024. Submitted for publication. `https://arxiv.org/abs/2410.08050`.

[22] Cliff C. Kerr, Robyn M. Stuart, Dina Mistry, Romesh G. Abeysuriya, Katherine Rosenfeld, Gregory R. Hart, Rafael C. Núñez, Jamie A. Cohen, Prashanth Selvaraj, Brittany Hagedorn, Lauren George, Michal Jastrzebski, Amanda S. Izzo, Greer Fowler, Anna Palmer, Dominic Delport, Nick Scott, Sherrie L. Kelly, Caroline S. Bennette, Bradley G. Wagner, Stewart T. Chang, Assaf P. Oron, Edward A. Wenger, Jasmina Panovska-Griffiths, Michael Famulare, and Daniel J. Klein. Covasim: An agent-based model of

COVID-19 dynamics and interventions. *PLOS Computational Biology*, 17(7):1–32, July 2021. Publisher: Public Library of Science.

[23] Pruthvi Kilaru, Dustin Hill, Kathryn Anderson, Mary B Collins, Hyatt Green, Brittany L Kmush, and David A Larsen. Wastewater Surveillance for Infectious Disease: A Systematic Review. *American Journal of Epidemiology*, 192(2):305–322, 10 2022.

[24] Martin Joachim Kühn, Daniel Abele, David Kerkmann, Sascha Alexander Korf, Henrik Zunker, Anna Clara Wendler, Julia Bicker, Dang Khoa Nguyen, René Schmieding, Lena Plötzke, Patrick Lenz, Maximilian Franz Betz, Carlotta Gerstein, Agatha Schmidt, Paul Johannssen, Margrit Klitz, Sebastian Binder, Martin Siggel, Wadim Koslow, Jan Kleinert, Kathrin Rack, Annette Lutz, and Michael Meyer-Hermann. Memilio v1.0.0 - a high performance modular epidemics simulation software, December 2023. `https://zenodo.org/records/10412635`.

[25] Martin J. Kühn, Daniel Abele, Tanmay Mitra, Wadim Koslow, Majid Abedi, Kathrin Rack, Martin Siggel, Sahamoddin Khailaie, Margrit Klitz, Sebastian Binder, Luca Spataro, Jonas Gilg, Jan Kleinert, Matthias Häberle, Lena Plötzke, Christoph D. Spinner, Melanie Stecher, Xiao Xiang Zhu, Achim Basermann, and Michael Meyer-Hermann. Assessment of effective mitigation and prediction of the spread of SARS-CoV-2 in Germany using demographic information and spatial resolution. *Mathematical Biosciences*, page 108648, 2021.

[26] Daniel B. Larremore, Bryan Wilder, Evan Lester, Soraya Shehata, James M. Burke, James A. Hay, Milind Tambe, Michael J. Mina, and Roy Parker. Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Science Advances*, 7(1):eabd5393, 2021.

[27] Ardashel Latsuzbaia, Malte Herold, Jean-Paul Bertemes, and Joël Mossong. Evolving social contact patterns during the COVID-19 crisis in Luxembourg. *PLOS ONE*, 15(8):1–13, 08 2020.

[28] H. Lau, T. Khosrawipour, P. Kocbach, H. Ichii, J. Bania, and V. Khosrawipour. Evaluating the massive underreporting and undertesting of covid-19 cases in multiple global epicenters. *Pulmonology*, 27(2):110–115, 2021.

[29] Lars Lorch, Heiner Kremer, William Trouleau, Stratis Tsirtsis, Aron Szanto, Bernhard Schölkopf, and Manuel Gomez-Rodriguez. Quantifying the effects of contact tracing, testing, and containment measures in the presence of infection hotspots. *ACM Trans. Spatial Algorithms Syst.*, 8(4), nov 2022.

[30] Louise A C Millard, Alba Fernández-Sanlés, Alice R Carter, Rachael A Hughes, Kate Tilling, Tim P Morris, Daniel Major-Smith, Gareth J Griffith, Gemma L Clayton, Emily

Kawabata, George Davey Smith, Deborah A Lawlor, and Maria Carolina Borges. Exploring the impact of selection bias in observational studies of COVID-19: a simulation study. *International Journal of Epidemiology*, 52(1):44–57, 12 2022.

[31] Alexander Mitranescu, Anna Uchaikina, Anna-Sonia Kau, Claudia Stange, Johannes Ho, Andreas Tiehm, Christian Wurzbacher, and Jörg E. Drewes. Wastewater-based epidemiology for SARS-CoV-2 biomarkers: Evaluation of normalization methods in small and large communities in southern germany. *ACS ES&T Water*, 2(12):2460–2470, Dec 2022.

[32] Shokoofeh Nourbakhsh, Aamir Fazil, Michael Li, Chand S. Mangat, Shelley W. Peterson, Jade Daigle, Stacie Langner, Jayson Shurgold, Patrick D'Aoust, Robert Delatolla, Elizabeth Mercier, Xiaoli Pang, Bonita E. Lee, Rebecca Stuart, Shinthuja Wijayasri, and David Champredon. A wastewater-based epidemic model for sars-cov-2 with application to three canadian cities. *Epidemics*, 39:100560, 2022.

[33] Matías Núñez, Nadia L. Barreiro, Rafael A. Barrio, and Christopher Rackauckas. Forecasting virus outbreaks with social media data via neural ordinary differential equations. *Scientific Reports*, 13(1):10870, 2023.

[34] Jordan Peccia, Alessandro Zulli, Doug E. Brackney, Nathan D. Grubaugh, Edward H. Kaplan, Arnau Casanovas-Massana, Albert I. Ko, Amyn A. Malik, Dennis Wang, Mike Wang, Joshua L. Warren, Daniel M. Weinberger, Wyatt Arnold, and Saad B. Omer. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nature Biotechnology*, 38(10):1164–1167, October 2020. Publisher: Nature Publishing Group.

[35] Elba Raimúndez, Erika Dudkin, Jakob Vanhoefer, Emad Alamoudi, Simon Merkt, Lara Fuhrmann, Fan Bai, and Jan Hasenauer. COVID-19 outbreak in Wuhan demonstrates the limitations of publicly available case numbers for epidemiological modeling. *Epidemics*, 34:100439, 2021.

[36] Andrew L. Rainey, Song Liang, Joseph H. Bisesi Jr., Tara Sabo-Attwood, and Anthony T. Maurelli. A multistate assessment of population normalization factors for wastewater-based epidemiology of covid-19. *PLoS ONE*, 18(4):e0284370, April 2023.

[37] Fachgebiet 32 Robert Koch-Institut. Abwassersurveillance AMELAG, September 2024.

[38] Manfred R. Schütze, David Butler, and M. Bruce Beck. *Development of the Integrated Simulation and Optimisation Tool SYNOPSIS*, pages 129–178. Springer London, London, 2002.

[39] Mark J. Siedner, Guy Harling, Zahra Reynolds, Rebecca F. Gilbert, Sebastien Haneuse, Atheendar S. Venkataramani, and Alexander C. Tsai. Social distancing to slow the US COVID-19 epidemic: Longitudinal pretest–posttest comparison group study. *PLOS Medicine*, 17(8):1–12, 08 2020.

[40] Statistisches Bundesamt. Mikrozensus. Accessible for 2020 to 2023 on `https://www-genesis.destatis.de/genesis/online?operation=statistic&levelindex=&levelid=&code=12211&option=table&info=on#abreadcrumb`.

[41] tandler.com GmbH. DYNA - Komplexes Parallelschrittverfahren - Verfahrensbeschreibung. Technical report, 2024.

[42] Deutscher Wetterdienst. Wetter und Klima - Deutscher Wetterdienst - Glossar - N - Niederschlagsintensität (dwd.de), 2024. `https://www.dwd.de/DE/service/lexikon/Functions/glossar.html?lv3=101906&lv2=101812` (visited on 02/20/2024).

[43] Fuqing Wu, Jianbo Zhang, Amy Xiao, Xiaoqiong Gu, Wei Lin Lee, Federica Armas, Kathryn Kauffman, William Hanage, Mariana Matus, Newsha Ghaeli, Noriko Endo, Claire Duvallet, Mathilde Poyet, Katya Moniz, Alex D. Washburne, Timothy B. Erickson, Peter R. Chai, Janelle Thompson, and Eric J. Alm. SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases. *mSystems*, 5(4):10.1128/msystems.00614–20, July 2020. Publisher: American Society for Microbiology.

[44] Shuxin Zhang, Elipsha Sharma, Ananda Tiwari, Yan Chen, Samendra P. Sherchan, Shuhong Gao, Xu Zhou, Jiahua Shi, and Guangming Jiang. The reduction of SARS-CoV-2 RNA concentration in the presence of sewer biofilms. *Water*, 15(11), June 2023.

[45] Henrik Zunker, René Schmieding, David Kerkmann, Alain Schengen, Sophie Diexer, Rafael Mikolajczyk, Michael Meyer-Hermann, and Martin J. Kühn. Novel travel time aware metapopulation models: A combination with multi-layer waning immunity to assess late-phase epidemic and endemic scenarios. 2024. Submitted for publication. `https://doi.org/10.1101/2024.03.01.24303602`.