TUM School of Computation, Information and Technology
Technical University of Munich

Modeling and simulation of SARS-CoV-2 transmission in dynamic crowds

Simon Andreas Rahn

Complete reprint of the dissertation approved by the TUM School of Computation, Information and Technology of the Technical University of Munich for the award of the

Doktor der Naturwissenschaften (Dr. rer. nat.).

Chair:

Prof. Dr. Stefanie Rinderle-Ma

Examiners:

1. Prof. Dr. Hans-Joachim Bungartz
2. Prof. Dr. Gerta Köster

The dissertation was submitted to the Technical University of Munich on 12 April 2024 and accepted by the TUM School of Computation, Information and Technology on 12 June 2024.

Abstract

The COVID-19 (coronavirus disease) pandemic, which raged worldwide between 2020 and 2023, has shown how seriously pathogens such as SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) can affect individuals and humanity as a whole. Simulation-based risk assessment constituted one approach to fight this crisis. It can give insights into human-to-human transmission during superspreading events, which governed the pandemic. However, there is no established model for estimating individual exposure to SARS-CoV-2. I address this gap by introducing an exposure model for pathogen transmission via inhalation of respiratory aerosols, implementing the model as reusable software, and running simulations with this tool. The exposure model describes how persons experience varying exposure risks while moving through space. It considers inhomogeneous pathogen concentrations in unventilated indoor environments. I use individual-based crowd models to capture human motion and extend them by aerosol clouds, which carry pathogens from infectious to susceptible persons. The representation of aerosols is deliberately simplified and matches the precision of the crowd model. I implement the exposure model as a flexibly adaptable module of an open-source simulation tool that enables me and others to analyze specific scenarios. Simulations of reenacted COVID-19 superspreading events help to build trust in the model. Based on this validation, I predict and compare the exposure risk for various situations where persons pass through indoor spaces. A global sensitivity analysis for these scenarios reveals that uncertainties related to the spread of aerosol clouds mainly influence output variability. This result suggests that future research efforts should be invested in studying the spread of aerosols to improve further predictions. Finally, I estimate exposure risks for different queuing situations. Comparing the scenarios indicates that crowd management could reduce exposure risks. Thus, I demonstrate how modeling and simulating pathogen transmission adds to quantitatively assessing exposure risks.

Zusammenfassung

Die COVID-19 (Coronavirus-Krankheit) Pandemie, die zwischen 2020 und 2023 weltweit grassierte, zeigte, wie schwerwiegend sich Pathogene wie SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) auf einzelne Personen sowie die gesamte Menschheit auswirken können. Die simulations-basierte Risikobeurteilung stellte einen Ansatz zur Bekämpfung der Pandemie dar. Diese kann aufschlussreiche Einblicke in die Mensch zu Mensch Übertragung bei Superspreading Events, die die Pandemie maßgeblich beeinflussten, gewähren. Allerdings gibt es kein etabliertes Modell zur Abschätzung der individuellen Exposition gegenüber SARS-CoV-2. Ich widme mich dieser Forschungslücke durch Einführung eines Expositionsmodells für die Übertragung von Krankheitserregern durch Inhalation respiratorischer Aerosole, dessen Implementierung als wiederverwendbare Software, sowie Durchführung von Simulationen mit diesem Werkzeug. Das Expositionsmodell beschreibt, wie Menschen variierenden Expositionsrisiken ausgesetzt sind, während sie sich im Raum bewegen. Es berücksichtigt inhomogen verteilte Pathogenkonzentrationen in ungelüfteten Innenräumen. Ich nutze individuen-basierte Personenstrommodelle, um menschliche Bewegung zu erfassen, und erweitere diese um Aerosolwolken, die Pathogene von infektiösen zu empfänglichen Personen transportieren. Die Darstellung von Aerosolen ist bewusst vereinfachend und passt zur Detailgenauigkeit des Personenstrommodells. Ich implementiere das Expositionsmodell als flexibel anpassbares Modul eines quelloffenen Simulationswerkzeuges, das es mir und anderen ermöglicht, spezifische Szenarien zu analysieren. Simulationen nachgestellter COVID-19 Superspreading Events helfen, Vertrauen in das Modell aufzubauen. Basierend auf dieser Validierung prognostiziere und vergleiche ich die Expositionsrisiken in verschiedenen Situationen, in denen Personen Innenräume passieren. Eine globale Sensitivitätsanalyse für diese Szenarien zeigt, dass vor allem Unsicherheiten in der Ausbreitung der Aerosolwolken die Variabilität des Outputs beeinflussen. Dieses Ergebnis legt nahe, dass weitere Forschungsanstrengungen in die Untersuchung der Ausbreitung von Aerosolen investiert werden sollten, um künftige Vorhersagen zu verbessern. Zuletzt schätze ich Expositionsrisiken in verschiedenen Warteschlangenszenarien ab. Der Szenariovergleich deutet daraufhin, dass Crowd-Management Expositionsrisiken reduzieren könnte. Auf diese Weise demonstriere ich, wie Modellierung und Simulation der Übertragung von Krankheitserregern zur quantitativen Bewertung von Expositionsrisiken beiträgt.

Plain language summary

The COVID-19 pandemic has shown how seriously the spread of a virus such as SARS-CoV-2 can affect our daily lives. To overcome the pandemic, it was important to find out how the virus jumps during so-called superspreading events from one ill person to many other healthy persons who are gathering in the same place. One can predict and study this virtually with the aid of mathematical descriptions of the real world. However, there is no useful description of how SARS-CoV-2 is carried from one person to another. Therefore, I define simple rules for how ill persons breathe out small particles that can carry the virus through the air over long periods of time and over long distances in a room. Other healthy persons can breathe in these particles. There is a high chance that persons who inhale many particles become ill. I also consider that humans move around while breathing in and out. Therefore, it can happen that there are more of these particles in one part of a room than in the other part. With these rules, I study how many healthy persons would breathe in a large number of the particles in situations of everyday life, such as going to a restaurant or to a choir practice. I compare my predictions to actual superspreading events and, thus, show that my rules seem correct. Then, I predict the risk for situations where people are waiting in a line, for example, in a supermarket. The results suggest that, on average, people would probably breathe in fewer particles if they wait in an orderly way. In this manner, I show how mathematical descriptions can help us separate risky from less risky situations so that we know what to do in reality.

Acknowledgment

This endeavor would not have been possible without my supervisors. I would like to express my gratitude to: Prof. Dr. Hans-Joachim Bungartz, especially for broadening my perspective on modeling and simulation and for always being available when I needed advice; Prof. Dr. Gerta Köster, for providing the foundation for my research activities at the Munich University of Applied Sciences HM, for her strong support and guidance.

I would like to extend my sincere thanks to Dr. Marion Gödel for accompanying me as a mentor and for our great collaboration. Special thanks go to my teammates of the HM Pedestrian Dynamics research group, particularly for mutual help and inspiration. I had the pleasure of discussing my research with partners and friends of the Pedestrian Dynamics group and peers in the field, whom I would like to thank for constructive and enlightening interchange.

Contents

List of figures

List of tables

Acronyms

2D	two-dimensional
3D	three-dimensional
ANOVA	analysis of variance
CFD	computational fluid dynamics
COVID-19	coronavirus disease
FAIR	findable, accessible, interoperable, and reusable
MERS	Middle East respiratory syndrome
OSM	optimal steps model
RNA	ribonucleic acid
SARS	severe acute respiratory syndrome
SARS-CoV	severe acute respiratory syndrome coronavirus
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SFM	social force model
SIR	susceptible-infected-removed
WHO	World Health Organization

1 Introduction

Pathogens are part of everyday life. They are found in the air, soil, water, and, of course, also in our bodies. Some cause infections with mild consequences, while other infections can end in death. The human body developed effective mechanisms to protect us against pathogens. However, sometimes the immune system cannot prevent infection. We become infectious ourselves and possibly pass the disease on to others. Despite substantial advances in disease prevention and control, infectious diseases regularly have regional to global socioeconomic repercussions. This problem comes up, for example, with the annual influenza wave circulating in our workplaces, schools, and other networks. It became dramatically apparent in 2020 when the outbreak of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) escalated into the coronavirus disease (COVID-19) pandemic. Three years later, the global situation subsided. However, zoonotic outbreaks, as was the case with COVID-19, appear to become more likely because of several factors: Climate change and land-use change aggravate the risk of zoonotic spillover [1]; Urbanization accelerates disease spread in a population because more and more people are crowded in one place [2]; Globalization catalyzes disease propagation into different parts of the world [2]. For these and other reasons, regional or global outbreaks of infectious diseases will continue to be a concern.

Fortunately, we can prepare ourselves against these threats. Disease prevention and control strategies, as well as monitoring and forecasting systems, undergo progress. Moreover, the COVID-19 pandemic demonstrated in an unparalleled way how collective scientific efforts can alleviate the situation. Large parts of the scientific community dedicated their workforce to fighting the global health crisis. Many researchers contributed through experiment and analysis, while others, particularly computational scientists, promoted modeling and simulation to predict and analyze fictional scenarios. Modeling and simulation can give insights into the transmission dynamics within a population and reveal the driving factors of a pandemic. This information is decisive for taking effective action against the spread and can support targeted crisis communication, for example, as upheld by Betz et al. [3]. Once a model has proven useful, it serves as a stepping stone for future challenges.

1.1 Motivation: on the lack of individual-based disease transmission models

The ability to tackle public health concerns increases with advancements in epidemiology. Among other things, epidemiology investigates the outbreak and the spread of infectious diseases in a population [4, p. 852]. In the case of COVID-19, it was found that the dynamics of the pandemic were strongly determined by superspreading events [5]. Such events occurred locally, for example, during mass gatherings or daily business,

as visualized in Fig. 1.1. This sketch illustrates what I call an *everyday situation*. To clarify this, I use the following definition throughout my dissertation:

Definition: everyday situation

An *everyday situation* is likely to happen in public spaces to an arbitrary group of people, for example, persons walking through a corridor or lining up in a queue and waiting for a particular service.

Understanding how critical such situations are in terms of pathogen transmission was of utmost importance during the COVID-19 pandemic for making informed decisions. However, lack of knowledge often impeded tailored actions and, instead, sometimes resulted in bizarre and debatable regulations while parts of the population ignored effective suggestions.

Figure 1.1: The sketch visualizes a situation of everyday life with a potential risk of pathogen transmission. Several persons, including one COVID-19 case (red), are queuing in line, for example, in a supermarket. Susceptible persons (blue) could become infected by inhaling respiratory aerosols that carry the virus (yellow). Note: own drawing.

Models that are capable of resolving inter-individual SARS-CoV-2 transmission at this scale would have been helpful in complementing laboratory and field research. However, the state of the art offers no such models. There is one class of models, based on the Wells-Riley equation [6, 7], that focuses on transmission via respiratory particles that persist in the air. This approach neglects effects, in particular, spatially varying exposure risks [8] or transmission via fomites and larger respiratory particles that sediment quickly [9]. In addition, the absence of calibration data makes it difficult for this class of models to calculate infection risks [9], particularly with regard to SARS-CoV-2 [10].

The other type of established small-scale models, such as the one proposed in [11], pursues an individual-based approach to account for transmission in dynamic human crowds. This concept appears promising, but it typically considers merely short-range transmis-

sion via larger respiratory droplets emitted, for example, when people cough or sneeze. Long-range transmission via pathogens that remain airborne for a certain time, which is important for SARS-CoV-2, is not included. Attempts toward combining individual-based approaches with long-range transmission include random human movement [12], movement patterns observed for scenarios in specific supermarkets [12, 13], in a certain hospital [14], and activities defined for a bar scenario [15]. However, random movement does not reflect real conditions, and the movement patterns used for specific scenarios are difficult to transfer to arbitrary everyday situations. Taken together, state-of-the-art models either do not resolve the required level of detail to get insights into inter-personal transmission, do not model long-range airborne transmission of SARS-CoV-2, or rely on unique activity schedules and movement patterns.

I am aware of only two recent developments toward long-range transmission in crowds that offer a basis for simulating realistic human movement in various situations. They were published around the same time or shortly after my initial proposals [16, 17]. Each of them focuses on distinct aspects and adopts different modeling approaches. Grignard et al. [18] extend an established model for crowd dynamics with transmission via proximity, homogeneously distributed airborne pathogens in a room, and fomites. However, the model introduces a wealth of parameters that are difficult to determine and handle. Mukherjee and Wadhwa [19] consider airborne transmission in an evacuation scenario in the presence of semi-analytically defined airflows. The simulated time is relatively short, which leads to exposure times of a few seconds, while critical exposure to SARS-CoV-2 is expected to require several minutes of close contact [20]. Increased efforts to model individual-based airborne transmission in recent years indicate that the problem of finding a suitable model has not yet been conclusively answered.

Once a mathematical model has been derived, it should also be used adequately. This means that numerical experiments should take into account uncertainties in the model parameters and their effect on the output. Many model parameters related to SARS-CoV-2 transmission are highly uncertain because we know too little to determine true values. Nonetheless, recently conducted simulation studies often consider only average inputs, which does not necessarily return reliable results. Understanding the consequences of possible variations in the model output is indispensable before assessing infection risks. In addition, investigating the model sensitivity with respect to parameter uncertainties should precede fine-tuning of a modeling approach, as stated in [21]. For these reasons, it is vital to scrutinize uncertainties in simulations.

Overall, the state of the art in modeling and simulation at small scales lacks an approach that represents the transmission characteristics of pathogens such as SARS-CoV-2. Recent developments attempting to fill this gap neglect uncertainties in the simulation. Consequently, it is unclear how much one can trust the simulation output and apply it in the real world.

1.2 Aims and research questions

The brief overview of modeling approaches demonstrated that there is no suitable model for simulating SARS-CoV-2 transmission between individuals. Therefore, I address the following question in this dissertation:

Research question

How can airborne transmission of pathogens such as SARS-CoV-2 be modeled and simulated?

This question involves sub-aspects

- about the theoretical background, state of the art, and recent developments:
  - What are the essential characteristics of SARS-CoV-2 transmission?
  - How does the literature from infectious disease modeling operationalize these mechanisms, and which facets are not adequately represented?
  - How should uncertainties about airborne pathogen transmission be addressed in modeling and simulation?
- about creating and implementing a model:
  - How can important aspects of inter-individual transmission of SARS-CoV-2 via airborne respiratory aerosols be described mathematically?
  - How can the mathematical model be implemented efficiently as sustainable software in the sense of reusable and verified code?
- about running simulations with the model and their practical relevance:
  - How can one reliably predict the exposure risk for everyday situations?
  - How can uncertainties in the simulation be quantified?

I seek to answer these questions by adopting an interdisciplinary approach. In this manner, I join in the call by Fefferman et al. for "a new paradigm for pandemic preparedness wherein interdisciplinary research and close collaboration with public policy and health practitioners can improve our ability to prevent, detect, and treat pandemics through tighter integration among domains, rapid and accurate integration, and translation of science to public policy, outreach and education, and improved venues and incentives for sustainable and robust interdisciplinary work" [22, p. 1]. On the one hand, I believe that the combination of and the interchange between disciplines opens up new perspectives and helps us think outside silos. On the other hand, interdisciplinary research entails a variety of obstacles. Above all, it demands an understanding of distinct disciplines, their respective methods and mindsets. It also means that one must establish a shared vocabulary. In this work, the word *agent* has great potential to confuse. Crowd researchers refer to virtual persons as agents, whereas, in microbiology, an infectious agent is a pathogen. I replace *agent-based modeling* with *individual-based modeling*, and *individual* refers to a single virtual person. Generally, I sometimes use longer descriptions

instead of standard terms to avoid other ambiguities or misleading framing. Thus, I aim to make my research comprehensible for a broad audience.

This goal of accessible research outputs applies not just literally to the terminology used in this dissertation but also figuratively to my modeling approach. I employ simple concepts that researchers across various disciplines should be able to understand, use, and extend to address their own subjects. To that end, I integrate my model into an established open-source simulation framework designed for research purposes. It offers interfaces for both developers who know how to modify and extend code and ordinary users who are content with setting up simulations using a graphical user interface. In this manner, I seek to foster interdisciplinary approaches in infectious disease modeling.

1.3 Structure of this work

The structure of this work mirrors the modeling cycle depicted in Fig. 1.2. This illustration deviates slightly from representations such as the simulation pipeline explained in [23, pp. 1–4] or the description in [24], but the concept remains the same. My first step is to observe the real world, followed by deriving and implementing a suitable model, to finally run simulations. This process is iterative and integrates improvements, for example, if the validation shows that the simulation results are a poor match for reality.

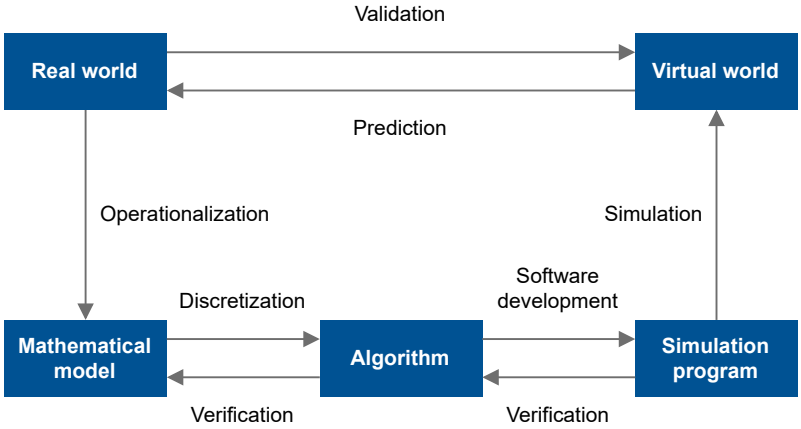


Figure 1.2: I adopt an iterative modeling approach in this work, starting with observing the real world.

To clearly set apart what is new in this dissertation, I divide it in two parts: Part I gives an overview of the findings discovered by others. It introduces the background knowledge required for deriving a transmission model for SARS-CoV-2 and for using it appropriately to obtain reliable outputs.

- Chapter 2 summarizes the scientific literature on SARS-CoV-2 transmission. This substitutes own experimental observations of the real world.

- Chapter 3 discusses how state-of-the-art models and recent developments operationalize these findings.
- Chapter 4 introduces the theory of uncertainty quantification methods.

Part II continues with my own contribution:

- Chapter 5 describes how I translate the findings from Chapter 2 into a mathematical model that aims to overcome deficiencies of approaches discussed in Chapter 3.
- Chapter 6 documents the implementation of my model and its verification. It also points out opportunities for extending the simulation tool.
- Chapter 7 is dedicated to validating the model and conducting numerical experiments to predict the exposure risk.
- Chapter 8 improves the reliability of these predictions with the aid of uncertainty quantification methods.

Finally, Chapter 9 summarizes my dissertation, evaluates the achievements, and provides an outlook on future directions. Appendix A gives details about the technical equipment and tools used in this work. The remaining appendices provide additional explanations and access to supplementary material referenced in the main part.

Part I

Background, state of the art, and recent developments

Modeling and simulation of airborne coronavirus transmission touches on various disciplines. The most important ones in the context of this work are computer science, mathematics, and life sciences. Since the art of modeling is in capturing the essence of the system under investigation, only a small fraction of each discipline is relevant for developing an adequate model. The following chapters of Part I discuss works by others on coronavirus transmission, epidemiological modeling, and uncertainty quantification. A subsection briefly introduces microscopic crowd simulation, as several individual-based epidemiological models are built on it. The knowledge and the number of scientific contributions about coronavirus transmission and epidemiological modeling experienced extraordinary growth from 2020 onward. Therefore, Part I not only contains background and state-of-the-art literature but also covers recent developments, that is, material published during the time of working on this dissertation from 2021 to 2024.

The rapid developments within these years have led to several challenges when reviewing the literature. Firstly, the continuous growth of knowledge necessitates an iterative process. Secondly, in addition to keeping track of several disciplines, the large amount of coronavirus-related contributions requires an efficient selection process. My search strategy in breadth involved screening databases, such as PubMed,[1] Elsevier's coronavirus information center,[2] and preprint servers (medRxiv[3] and arXiv[4]). This rather broad screening was complemented by references mentioned in the podcast Coronavirus-Update [25], in which two of Germany's leading virologists, Sandra Ciesek and Christian Drosten, discussed the latest scientific findings. The Robert Koch-Institut [26] regularly published updates of an epidemiological bulletin, offering further reading, while Covid Reference[5] delivered summaries of essential research outputs. I also used personalized recommendations for publications through Research Gate.[6] The in-depth search narrowed down the results listed by Scopus[7] and Google Scholar.[8] Tracking the references of publications either through the list of references or through CrossRef's[9] cited-by service revealed connections between publications.

Facing the extensive efforts of the scientific workforce responding to the outbreak of the coronavirus disease, I have certainly missed some information despite this intensive literature review. Considering that research on the coronavirus is ongoing and that this work takes the coronavirus as an example of any airborne pathogen, simplifying some developments and preferring systematic review studies over primary sources seems justified.

[1] www.pubmed.ncbi.nlm.nih.gov
[2] www.elsevier.com/connect/coronavirus-information-center
[3] www.medrxiv.org
[4] www.arxiv.org
[5] www.covidreference.com
[6] www.researchgate.net
[7] www.scopus.com
[8] https://scholar.google.de
[9] www.crossref.org
All URLs accessed on April 2, 2024

2 Transmission of the severe acute respiratory syndrome coronavirus 2

On December 31, 2019, the World Health Organization (WHO) received information about several cases of pneumonia in Wuhan, China. A novel virus responsible for this illness was identified a few days later [27]. It was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [28] and determined to cause the coronavirus disease (COVID-19) [29]. Within the following weeks, the outbreak rapidly evolved into an epidemic, which first affected larger geographical areas. Finally, the disease spread across the world. In response to this global public health emergency, the WHO [30] declared COVID-19 as a pandemic on March 11, 2020. Governments advised regulations that deeply intervened with everyday life, such as lockdowns, stay-at-home orders, and social distancing. Thus, they tried to contain the spread and prevent healthcare systems from being overwhelmed. Three years later, as of May 5, 2023, over 764.6 million laboratory-confirmed cases of COVID-19, including more than 6.9 million deaths, have been reported to the WHO [31]. However, immunity increased as the number of infections rose and vaccines became available. This globally alleviated the situation. According to the WHO Director-General, COVID-19 has reached the status of a permanent health problem and can no longer be regarded as a so-called public health emergency of international concern [32]. Science significantly contributed to overcoming the crisis by identifying the root cause, deciphering the virus, figuring out dominant transmission paths, evaluating the effectiveness of containment strategies, and finally developing effective vaccines. The following sections cover several of these aspects that are relevant to my research, including how the massive mobilization of scientific brainpower has temporarily changed scientific procedures and how this influences the present work.

2.1 Properties of the virus and how it spreads

This section summarizes the features of SARS-CoV-2 and its transmission modes. The explanations focus on a few selected aspects, providing a knowledge base sufficiently profound to understand the subsequent chapters about mathematical modeling of infectious diseases and, particularly, how to translate SARS-CoV-2 transmission into an algorithm.

2.1.1 Fundamental characteristics of SARS-CoV-2 and COVID-19

SARS-CoV-2 is the virus causing COVID-19. It belongs to the species *Severe acute respiratory syndrome-related coronavirus*[1] and is placed within the family *Coronaviridae*[1][28].

[1] Following the International Code of Virus Classification and Nomenclature, the family and species are italicized, and the first letter is capitalized, whereas the name of the virus is to be written in standard script and, unless part of the name is a proper noun, lower case [33].

Members of the *Coronaviridae* are enveloped viruses that use ribonucleic acid (RNA) to encode their genetic information [34]. The name corona is borrowed from the Latin word for crown. It originates from the appearance of the virus particle when located outside of a cell, also called a virion. A spherical morphology with spike proteins projecting from the surface of the virion (see Fig. 2.1) resembles a solar corona [35]. Coronaviruses affect many animals [36] and, in certain cases, also humans [37]. The outbreak of COVID-19 is the result of cross-species transmission of an animal coronavirus to humans [28]. This is the third documented zoonotic emergence of coronaviruses since 2002 leading to severe diseases in humans. While the outbreaks of the severe acute respiratory syndrome (SARS) in 2002 to 2003 and the Middle East respiratory syndrome (MERS) in 2012 resulted in major epidemics, COVID-19 escalated into a pandemic [38].

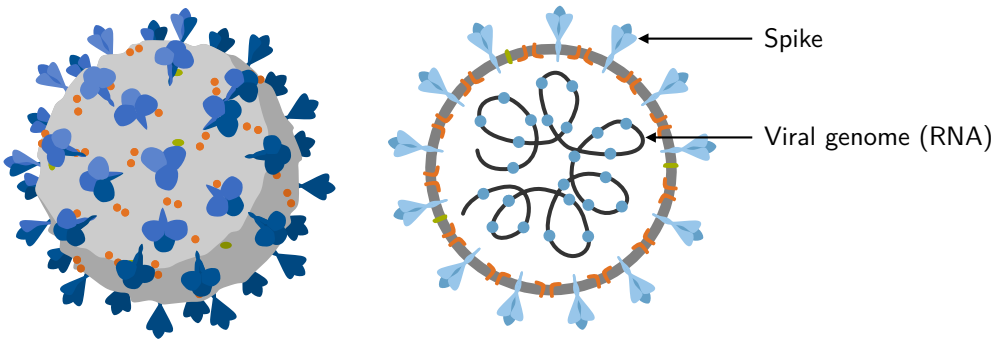


Figure 2.1: Schematic representation of the structure of the human coronavirus. Note: own figure adapted from [39].

The more a virus circulates in a population, the more it can change as a virus constantly replicates. The replication machinery introduces small changes in the copies, so-called mutations, and a new variant of the virus evolves when enough mutations accumulate. In the case of SARS-CoV-2, this led to, for example, altered transmissibility, disease severity, or other epidemiological or clinical properties in the especially concerning variants Alpha, Beta, Gamma, Delta, and Omicron. Omicron showed the highest transmissibility among these [40, 41], while higher morbidity and mortality were found in cases infected with the Delta variant [42].

Generally, COVID-19 can cause a variety of clinical manifestations. They range from asymptomatic to severe symptoms, which can end up in multiple organ failure and death [41]. Da Rosa Mesquita et al. [43] systematically review studies published during the early phase of the pandemic and find fever and cough to be the most and second most common symptoms, respectively. Cough is of particular interest because it is connected to the major transmission pathway through respiratory droplets [43]. Even without showing symptoms, also pre-symptomatic cases or, despite lower average infectiousness, asymptomatic cases can infect others [44].

However, the infection process is not fully understood. For example, the amount of viruses required to successfully start an infection is unclear. In general, immunity to pathogens decreases with increasing age [45]. The human challenge study conducted by Killingley et al. [46] addresses the immune response in young adults after intranasal administration of the virus. A relatively low dose of a wild-type SARS-CoV-2 strain caused infections in 53 % (18) of the participants who had neither been vaccinated nor infected prior to inoculation. This study does not represent an average population because of a sampling bias and a small sample size. Furthermore, human challenge trials involve ethical issues by nature. A review study [47] finds a link between the dose and infection in humans, but the infectious dose in humans remains uncertain. Karimzadeh et al. [48] draw a similar conclusion based on an analysis of animal experiments and literature on the infectious dose in humans for several other viruses. They suggest that SARS-CoV-2 is more contagious than influenza viruses. According to Popa et al. [49], the transmission bottleneck size, which represents the number of viral particles that start an infection, is, on average, 10^3 particles. Martin and Koelle [50] re-examine the data provided by Popa et al. and argue that the transmission bottleneck is much smaller than the primary investigation suggests. Overall, the review studies [47, 48] and inconsistent findings in [49, 50] indicate that there is no consensus on the number of viruses required for infection, which impedes a quantitative dose-response assessment.

Once an infection has started within an individual, the viral load can reach 10^9 to 10^{11} virions during the peak of infection [51]. Infectious virus titers and RNA levels reach their maximum around the day of symptom onset [52]. Killingley et al. [46] report the highest concentrations of viral copies in a patient's nose. SARS-CoV-2 reproduces not only but very efficiently in the nose, which explains the strong role of shedding via the upper respiratory tract [53]. It should be noted that the emission rates are highly individual, vary throughout an infection, and depend on respiratory activity and other factors. Fig. 2.2 exemplary visualizes exhalation rates observed under various conditions. Malik and Kunze [54] analyze the exhaled breath of two patients over the course of infection. At the peak of infection, the exhalation rates for normal breathing reached $4.1 \cdot 10^4$ and $2 \cdot 10^5$ RNA copies per hour. Lai et al. [55] collected the exhaled breath aerosol of patients infected with different variants of SARS-CoV-2. The participants provided 30-minute breath samples according to a protocol that includes loud speaking, singing, coughing, and sneezing, as described in [56]. The viral loads ranged from below the limit of detection (75 RNA copies per sample) to $1.8 \cdot 10^7$ copies per sample. Interestingly, the fraction of smaller aerosol particles contained higher viral loads than the larger particles. Alsved et al. [57] document orders of magnitude lower values for individuals infected with Alpha or pre-Alpha variants. The observed median emission rates reach 70, 110, and 80 RNA copies per minute for breathing, talking, and singing, respectively. One person produced $7.8 \cdot 10^3$ RNA copies per minute during singing.

These findings indicate that SARS-CoV-2 transmission is closely linked to transmission via exhalation and inhalation of aerosolized respiratory fluids. The smaller an exhaled particle, the lower the chance it contains a virion. Obviously, the particle must be larger than the virion. The size of a SARS-CoV-2 virion ranges from 60 nm to 140 nm with

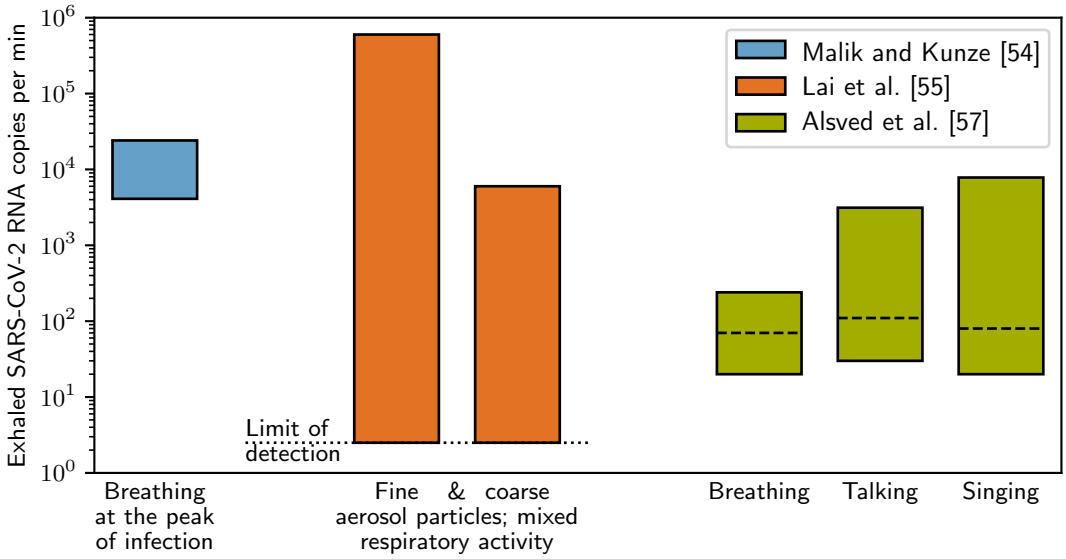


Figure 2.2: The diagram compares exhalation rates (SARS-CoV-2 RNA copies per minute) reported for different conditions. Lai et al. [55] distinguish between exhalation rates attributed to fine ($\leq 5\,\mu m$) and coarse ($> 5\,\mu m$) aerosol particles collected for mixed respiratory activity, including loud speaking, singing, coughing, and sneezing. Boxes span the range of data. The dashed lines represent median values where available.

spikes of 9 nm to 12 nm length [58]. Therefore, the minimum diameter of a particle carrying SARS-CoV-2 is approximately 0.1 µm [59]. Cvitešić Kušan et al. [60] collected genetic material in different indoor hospital and home care environments where infectious persons were present. The authors analyzed particle sizes between 0.01 µm to 32 µm and found RNA copies mostly in particles greater than 0.18 µm. Smaller particles probably carry only fragmented SARS-CoV-2. Thus, human respiration promotes viral shedding.

In summary, SARS-CoV-2 is an infectious RNA virus that is capable of infecting and causing severe damage in humans. Neither infective dose nor viral shedding have conclusively been analyzed, but they could vary in the order of magnitudes. Typical for respiratory viruses, it is transmitted through exhalation and inhalation. The following section discusses the transmission modes of SARS-CoV-2 in more detail.

2.1.2 Transmission paths

The term *transmission* has a variety of connotations. In the context of infectious diseases, it refers to the passing of a pathogen from an infected host to a susceptible individual. The mode of transmission is typically linked to the dominant habitat within the host. For example, respiratory pathogens usually take the airborne route, while intestinal pathogens are carried by vehicles contaminated with feces [4, p. 855]. Some pathogens have more than just one transmission path. In such a case, the infectivity may vary de-

pending on which path the pathogen chooses [61]. For example, infection via respiration often needs fewer pathogens than ingestion [62, p. 45]. Furthermore, each transmission path for the same pathogen can also change the degree of damage caused by the disease. When the probability or severity of infection depends on the transmission path, this can be described as an anisotropic infection [63].

In the scientific literature and the public discourse, one is often confronted with a traditional classification of pathogens according to their transmission mechanisms. Fig. 2.3 divides these mechanisms into direct and indirect transmission, as described in [4, p. 856]. Direct transmission occurs without the aid of intermediate hosts or inanimate objects. This mechanism requires physical contact or droplets emitted through respiratory activities, for example, coughing or sneezing. Typically, droplets are defined as particles with an aerodynamic diameter of $> 5\,\mu m$. It is assumed that these relatively large droplets reach only a short distance of 1 m to 2 m [64], hence the terms droplet transmission and short-range transmission are used as synonyms. Within this range, pathogens contained by the droplets can directly deposit onto another person's skin or mucous membranes and, eventually, enter the body. Smaller respiratory droplets, typically with a size below $5\,\mu m$ and often termed droplet nuclei, are associated with the direct route but sometimes also counted among the indirect route. They remain airborne and can carry pathogens for several minutes to hours over a longer distance [65]. Indirect transmission is facilitated by living carriers, so-called vectors, or non-living objects. Non-living objects that transport viable pathogens are often referred to as fomites or vehicles. Examples of fomites in daily life are door handles and handrails. Vehicles are usually objects that affect many individuals, as is the case for water and food [4, p. 856]. Another common characterization distinguishes contact, droplet, and airborne transmission [65].

The conventional classification of direct and indirect or droplet and airborne transmission can be misleading. In search of a physically more accurate description, Jones and Brosseau [66] develop the concept of aerosol transmission. However, this definition has neither replaced the classical categorization, nor is it being used consequently throughout the literature. Aerosol transmission is sometimes meant to involve both droplet and airborne transmission, and sometimes it is used interchangeably only for airborne transmission [67]. In addition, the threshold size to distinguish between droplets and airborne respiratory particles is unclear. Typically, it is set to $5\,\mu m$, although this seems an inappropriate mark for the transition between two significantly different aerodynamic behaviors. Tellier et al. [67] propose $10\,\mu m$, while other authors argue for $100\,\mu m$ instead [64].

Following the conventional definitions, there is convincing evidence that human-to-human transmission of SARS-CoV-2 mainly occurs via droplets and airborne respiratory particles [64, 68–70]. It is important to recognize that there has been a debate about the relative contributions of each transmission path to the dynamics of the pandemic. Despite the scientific evidence, the role of airborne transmission had not been recognized for several months after the outbreak of COVID-19 [68].

One argument for droplet and airborne transmission of SARS-CoV-2 is that other respiratory viruses, such as its genetically similar forerunner severe acute respiratory

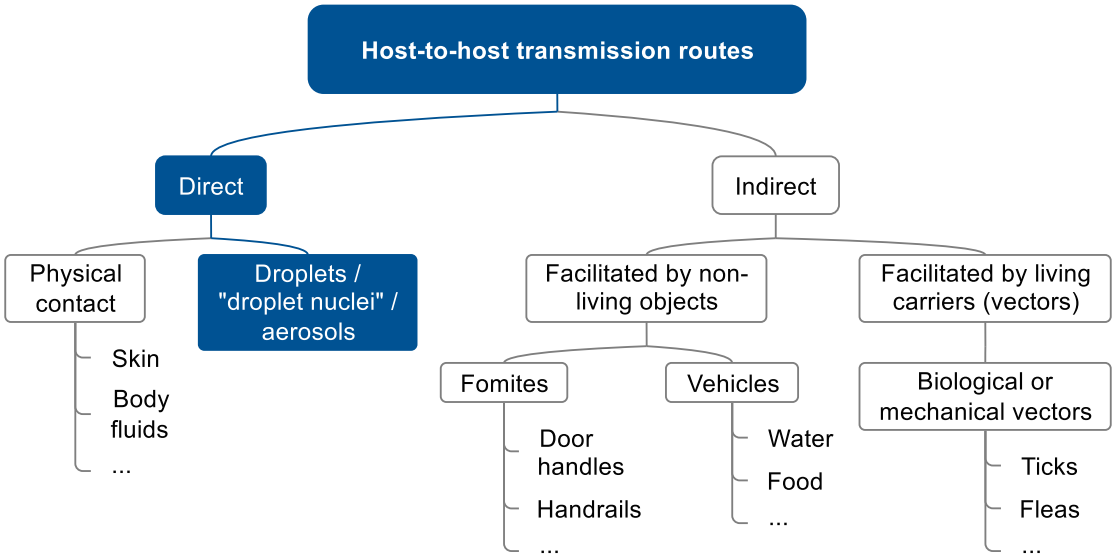


Figure 2.3: The diagram categorizes pathogen transmission routes according to a commonly used classification. The highlighted transmission routes dominate the dynamics of the COVID-19 pandemic. However, there is no clear distinction between large droplets and fine aerosol particles, which makes the definition of droplet and airborne transmission ambiguous.

syndrome coronavirus (SARS-CoV), MERS coronavirus, measles virus, influenza virus, and many more employ these mechanisms [64]. Analogical reasoning leads to the hypothesis that SARS-CoV-2 exhibits comparable transmission mechanisms. This is also supported by several studies.

Most importantly, retrospective analyses of local outbreaks of COVID-19 associated with specific events strongly suggest that airborne transmission represents a dominant mechanism under certain conditions, particularly in indoor environments with inadequate air exchange. The infection patterns observed for two clusters recorded early in the pandemic, a choir rehearsal in Skagit Valley, Washington, USA [71, 72] and in a restaurant in Guangzhou, China [72–74], cannot be explained without assuming airborne transmission. The choir rehearsal and the restaurant setting are just two of a myriad of superspreading events, some of which are listed in [75]. They shaped the understanding of the transmission dynamics of the COVID-19 pandemic [76]. A superspreading event, as defined by Lloyd-Smith et al. [77], is an event where one or more individuals infect an extraordinarily high number of other persons. It is not an exception that infectious diseases are transmitted by 20 % to 80 % of the population [78]. Also, SARS-CoV-2 exhibits heterogeneous transmission dynamics. It is even possible that 10 % to 20 % infect 80 % of the population under certain circumstances [5]. Althouse et al. [5] explore which conditions facilitate superspreading events and identify closed environments, inadequate ventilation, crowded areas, and long exposure times as risk factors. Unfiltered and recirculated air can also contribute to transmission indoors [75]. These findings indicate

airborne transmission because larger droplets are not affected by ventilation. For this reason and since SARS-CoV-2 seems to be transmitted rather indoors than outdoors, Wang et al. [64] argue that airborne transmission is more dominant than droplets or fomites.

The primary role of the airborne route derived from superspreading events is in line with observations of the trend of the pandemic. Zhang et al. [69] compare the number of recorded COVID-19 cases in epicenters where different transmission mitigation strategies were implemented. By establishing causality between the mitigation strategies and the observed trend of infections, the authors derive that airborne transmission is the dominant mode. However, the actual causal strength is difficult to estimate. The fact that rules for preventing the spread have been declared does not necessarily mean that the population complies with these rules. Furthermore, other effects, such as anticipation of mitigation strategies by the population or undetected cases, could distort the results.

Collectively, these studies outline the critical role of mitigation strategies, ambient conditions, whether one considers indoor or outdoor spaces, and many other factors for the transmission characteristics of SARS-CoV-2. This complicates the quantification of the relative contributions of droplet and airborne transmission to the dynamics of the pandemic. To my knowledge, this has not yet been achieved. Nonetheless, the arguments above strengthen the hypothesis that droplet and airborne transmission play a crucial role and that airborne transmission is associated with indoor spaces, while droplet transmission could prevail outdoors.

The possibility of fomite transmission cannot be excluded because SARS-CoV-2 can remain viable for days on surfaces under favorable conditions [79, 80]. However, these studies have been criticized because they presuppose unrealistic conditions [81]. Harvey et al. [82] conclude from a longitudinal study and quantitative microbial risk assessment that fomites contribute minimally to community transmission of SARS-CoV-2. Based on this work and critically reviewing several studies on the stability of SARS-CoV-2 and SARS-CoV deposited on surfaces, Goldman [83] argues that fomite transmission plays only a subordinate role and should be treated accordingly with less effort. Referring to further studies, Mondelli et al. [84] endorse Goldman's [81] argumentation. Other transmission media, for example, environmental fecal wastes, wastewater, and water exposure are less relevant [85].

To conclude this section, a growing body of literature identifies airborne and droplet transmission as governing transmission mechanisms for the COVID-19 pandemic. The initial opinion that droplet transmission is the most crucial mechanism has been refuted, and there are voices advocating airborne transmission instead. The following section provides the background knowledge on why this transmission pathway is relevant.

2.1.3 The role of aerosols in coronavirus transmission

Respiratory aerosol particles, small and large, contributed to the spread of COVID-19. This section provides insight into the physics behind transmission via aerosols. For further reading, I refer to the textbook by Hinds [86], which breaks down aerosol science in theory and practice in an accessible way.

Prior to delving deeper into the topic, a few physical definitions are necessary to avoid ambiguities:

- An **aerosol** is composed of particles suspended in a gas, normally air [86, p. 4].
- The **aerosol particles**, also referred to as particles in the following, can either be solid or liquid [86, p. 4]. In the context of this work, they are aqueous droplets.
- An important quantity is the **particle size**. It influences which law of nature determines the behavior of the particles and, thus, affects the property of an aerosol significantly. Particles with diameters ranging from 0.002 µm to over 100 µm are usually considered aerosol particles [86, p. 8]. Since aerosols are often polydisperse, consisting of particles with different sizes, they are commonly characterized by a distribution of diameters.
- The diameter typically represents the **aerodynamic diameter**, which is the diameter of a spherical particle with a density of $1000\,\mathrm{kg\,m^{-3}}$ with an equal settling velocity, that is, the fall velocity of a particle experiencing gravitation in still air, as the considered particle [86, p. 53].

Aerosol particles are not infectious per se, but they can carry infectious pathogens that are hazardous to health. Smaller aerosol particles are of particular importance because they can carry a large fraction of emitted SARS-CoV-2 virions, keep them airborne for a long period, and may reach areas in the human respiratory system where the virus leads to severe damage [87]. The three related aspects of aerosol generation, the fate of aerosols after exhalation, and deposition in the human lungs are covered in the following.

2.1.3.1 Aerosol generation

Besides human respiration, there are many processes that generate pathogen-carrying aerosols. For example, aerosol particles can be resuspended from surfaces [88] or originate from medical procedures [89, 90], and there is aerosolization of wastewater when flushing the toilet [91]. However, these very specific circumstances are beyond the scope of the present work. The focus is on human respiration.

The physics of aerosol generation in the human respiratory system has not yet been completely uncovered because it involves complex processes, which can be summarized as follows: Human airways are covered with a fluid film. In the case of an infectious person, this film contains virions. Respiratory activities, such as coughing or breathing, aerosolize parts of the surface film and emit the aerosol particles into the environment. Morawska et al. [59] describe in depth the physical mechanisms of turbulent aerosolization and fluid film, filament, or bubble breakage.

The resulting aerosols have varying concentrations and particle size distributions. These characteristics determine the time certain amounts of pathogen remain airborne or how quickly they sediment. Over the past decades, several experimental studies, for example, [92–97] analyzed the size distributions for respiratory events but came to different results. Of course, the measurement techniques have changed, as other instruments

have become available since early investigations by Duguid [92] in 1945. Furthermore, the experiments were conducted under different conditions and had a specific objective, making comparisons between the results difficult. Clearly, the experiment setup and evaluation varied, including the instructions to produce aerosols, the composition of participants, and correction of measurements for particle shrinkage. For a detailed review and synthesis of available data, I refer to the work by Pöhlker et al. [98].

Based on the preprint version of this review [99], Bagheri et al. [97] carefully carry out a comprehensive experiment with 132 participants, aged between 5 and 80 years, in a controlled environment. Thus, they avoid many deficiencies of previous studies. As summarized in Fig. 2.4, vocalization leads to both smaller and larger particles, while breathing mostly produces particles with diameters below 5 µm. The number concentrations[2] for particles < 5 µm by respiratory activity decrease in the following order: shouting, singing, speaking, and breathing. For particles < 5 µm, the concentration increases with age, but the influence of other factors is negligible. One limitation is that the subjects were healthy [97]. Persons with respiratory infections can produce significantly more aerosols [100].

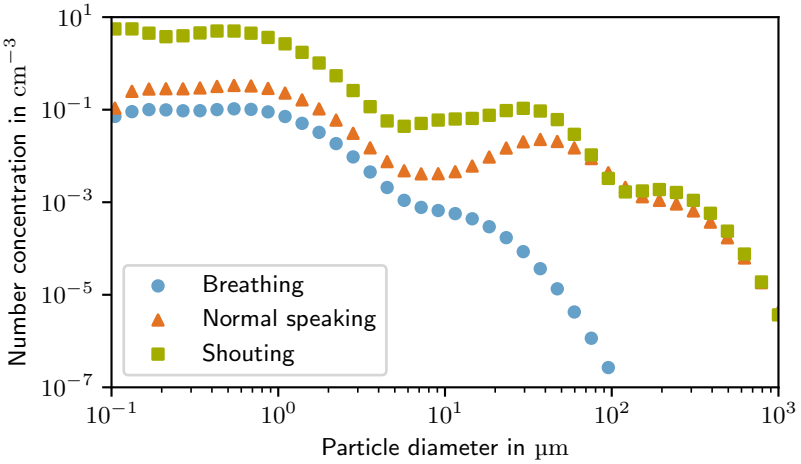


Figure 2.4: The particle size distribution depends on the respiratory activity. Note: own representation reproduced from [97], data retrieved from `https://aerosol.ds.mpg.de/en/` on September 5, 2023.

In an experimental setting, breathing may appear as the least crucial respiratory activity because it is associated with lower aerosol concentrations. In reality, breathing is highly relevant for pathogen transmission via aerosols, maybe even the deciding aspect. Shouting is atypical for everyday situations, talking can sometimes be avoided, and, if the emitter is symptomatic at all, coughing or sneezing are sporadic respiratory events. In contrast, breathing is a life-sustaining and continuous process, leading to a throughput

[2] The number concentration is also referred to as number density. Here, it measures the number of aerosol particles per unit volume.

of $10\,\mathrm{m}^3$ to $25\,\mathrm{m}^3$ of inhaled and exhaled air per day [86, p. 233]. The amount of shared indoor air and, hence, the number of inhaled aerosol particles originating from breathing can be high.

The aerosol number concentration and particle size distributions can act as indicators for viral load, but the relationship is not linear. It has been demonstrated for the influenza virus that breathing yields higher pathogen concentrations in the fraction of aerosol particles $< 5\,\mu\mathrm{m}$ [101]. Recent experiments confirm the same for SARS-CoV-2 [57, 102]. More precisely, particles with a size of $0.94\,\mu\mathrm{m}$ to $2.8\,\mu\mathrm{m}$ contain the largest portion of SARS-CoV-2 RNA, and 90 % of the viral RNA is found in particles with a dry size of less than $4.5\,\mu\mathrm{m}$. However, accounting for evaporation, the initial diameter could be larger by a factor of up to five [102]. The next subsection scrutinizes evaporation and other effects on the persistence of airborne pathogens.

2.1.3.2 Fate of exhaled aerosol particles and their viral load

The fate of environmental contaminants, in the present case airborne pathogens, describes what happens to substances once they have been released into the environment. The spatiotemporal spread of respiratory aerosols after exhalation is governed by interdependent physical laws, above all, gravitation, inertia, and evaporation [6, pp. 8–19]. To reduce the complexity, one often considers a single particle of a certain size and ignores that particles within a particle cloud actually have different properties than a single particle in air. For example, the relative humidity within an exhaled jet can be higher than in ambient air [103]. Hence, assuming either the ambient relative humidity around a single particle or the relative humidity of the puff results in different evaporation rates and settling velocities [104]. Such limitations must be kept in mind when analyzing the properties of an aerosol.

In general, the motion and evaporation properties of aerosol particles can be summarized as follows: For larger particles, gravity is more dominant than evaporation and vice versa for smaller particles (see Fig. 2.5). While larger respiratory droplets follow semiballistic trajectories and reach the ground within a few seconds, smaller aerosol particles are caught in the jet of exhaled air. Once the coherence of the jet has broken down, ambient airflows determine the particle motion [59]. It depends mostly on the ambient conditions where to draw the line between small and large particles [105]. The medical literature differentiates between aerosol particles and droplets with diameters smaller and larger than $5\,\mu\mathrm{m}$, respectively. However, this classification is misleading because, firstly, the term droplet suggests that these particles are not aerosol particles, which is not true. Secondly, the rather arbitrary threshold of $5\,\mu\mathrm{m}$ is ill-suited to capture the continuous transition of the dominant transport mechanism from smaller to larger particles [65].

Fig. 2.6 underlines the scales on which different physical processes play a role. The smaller aerosol particles deposit slowly. For instance, a $1\,\mu\mathrm{m}$ particle in still air with velocity $v \approx 3.5 \cdot 10^{-5}\,\mathrm{m\,s^{-1}}$ would settle roughly $vt = 2\,\mathrm{mm}$ within a time frame of $t = 60\,\mathrm{s}$. On such small scales, for example, inside the human lung, thermal diffusion is relevant [86, p. 150]. However, on the scale of a room, the dynamics of particles are determined by turbulent dispersion [59]. This means that airflows often disrupt and,

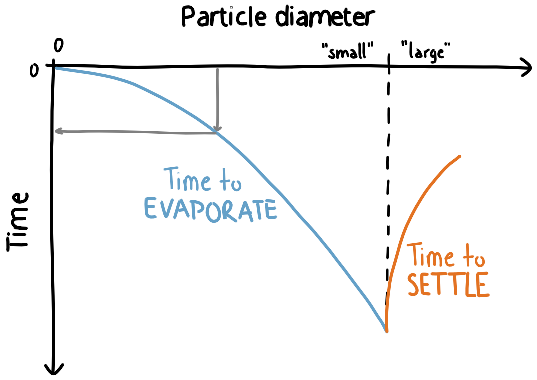


Figure 2.5: The Wells evaporation falling curve shows qualitatively how evaporation and settling affect aqueous droplets with different diameters. Larger particles settle faster than they evaporate, whereas smaller particles evaporate before they reach the ground. Note: own figure adapted from [105].

thus, prolong the settling of airborne particles. Airflows are caused by, for example, ventilation or humans dissipating thermal energy [106].

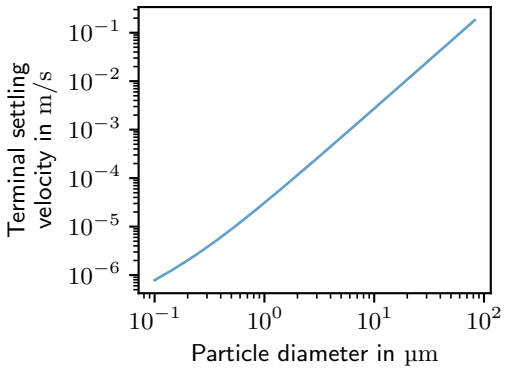


Figure 2.6: The curve shows the terminal settling velocity as a function of particle diameter for a water droplet settling in still air. The settling velocity results from balancing drag force and gravitational force (see Eq. 3.21–2 in [86, p. 49]). I calculated it for standard density spheres at 293 K (20 °C) and an atmospheric pressure of $101 \cdot 10^3\,\mathrm{N\,m^{-2}}$. Note: own figure.

They also affect the transport of fine aerosol particles in the horizontal direction. Already exhaling aerosol particles involves air currents, which carry aerosol particles over moderate distances. Hossain and Faisal [107] simulate a steady state air jet flow produced by a constantly breathing human. The results show that 1 µm aerosol particles travel about 0.3 m in the horizontal direction. Xie et al. [105] find similar horizontal displacements for normal breathing. The study includes the impact of varying particle sizes, emission velocities, mouth-opening sizes, and evaporation under varying ambient

conditions. Aerosol particles with diameters up to 200 µm are calculated to travel less than 1 m [105]. Kudryashova et al. [87] investigate the propagation of aerosols in an unventilated environment. They artificially generate aerosols with sizes between 0.5 µm and 20 µm. Within 3 min to 5 min, the aerosol concentration levels off at 1.5 m distance around the aerosol generator.

Importantly, evaporation also affects particle transport and sedimentation. Depending on the ambient conditions, the aqueous part of the particle evaporates until the diameter reaches stability. The remaining droplet nucleus contains non-volatile substances, such as mucins or proteins. The equilibrium diameter is, on average, around 1 µm [59]. Such small diameters are associated with a higher lifetime in the air. The time to reach the stable size ranges from milliseconds to seconds [108]. However, a high relative humidity can also lead to the growth of particles through condensation. Larger particle sizes lead to faster sedimentation [105]. As a consequence of evaporation and condensation, the particle diameter is a highly dynamic quantity. Therefore, it is difficult to characterize particles by their diameter [65]. Nevertheless, this quantity is frequently used and often considered constant for simplicity.

Sedimentation and evaporation have an impact on the amount of viable virus carried by aerosol particles. Chatterjee et al. [109] demonstrate that there is a correlation between virus survival time and the lifetime of aerosol particles. Employing a semi-analytical model, they estimate that the virus is inactivated roughly within 3 h in smaller particles (< 5 µm). An experiment for particles with the same upper size limit shows a similar timescale, with virus titer decreasing exponentially at a median half-life of about 1.1 h [79]. Smither et al. [110] analyze the stability of SARS-CoV-2 in artificial saliva with particle sizes between 1 µm to 3 µm under different conditions. Depending on the relative humidity, the half-life reaches values between 30 min to 170 min. A much shorter half-life of 5.5 min to 7 min is determined for number concentrations produced by coughing or sneezing [111]. Coughing or sneezing also produces larger particles [95]. Since larger particles sediment faster, a shorter half-life seems reasonable. If ventilation is involved, the half-life can be much shorter even for smaller particles. An experimental study demonstrates that the number concentration of aerosols with an average particle size of 5 µm in a ventilated room with a door and window opened can half within 30 s [112]. The results of these experiments and modeling studies attempting to determine the persistence of viable SARS-CoV-2 in aerosols must be interpreted carefully. Each survey focuses on specific conditions that do not directly reflect reality. Deriving an exposure or infection risk requires even more information about how airborne pathogens enter susceptible individuals.

2.1.3.3 Respiratory deposition

The final step in the transmission of pathogens via aerosol is the intake of a particle. Large respiratory droplets, well beyond 100 µm, that follow ballistic trajectories are less affected by air currents than smaller aerosol particles. Therefore, they enter the respiratory tract less frequently but rather deposit directly onto mucus membranes, such as the eyes [65]. In contrast, smaller particles can be inhaled and, thus, reach different regions

within the respiratory system [86, pp. 233–47]. The fraction of particles that enters the respiratory system through the nose or mouth compared to the particle concentration in the environment is defined as the inhalable fraction (see Fig. 2.7). It depends on the aerodynamic diameter and the flow field of the ambient air [86, p. 246].

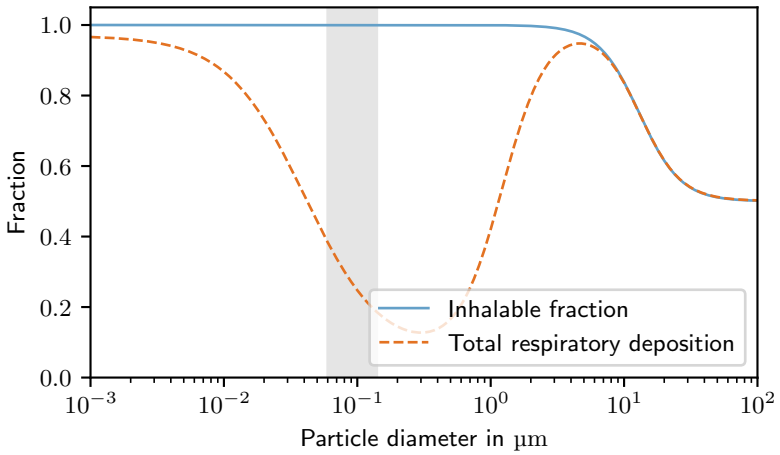


Figure 2.7: The diagram estimates the inhalable fraction and total deposition in the respiration system based on Hinds's adaption (Eq. 11.1–5 in [86, pp. 244–5]) of the International Commission on Radiological Protection [113] deposition model. The inhalable fraction is averaged over all directions of the airflow passing by an inhaling person. The deposition is fitted to averaged data from males and females as well as different levels of physical activity. The shaded area indicates the size range of SARS-CoV-2 virions [58]. Note: own figure.

Once aerosol particles have been inhaled, a part deposits in the respiratory tract or is exhaled again. Respiratory deposition involves several physical processes, primarily impaction, settling, and diffusion [86, pp. 235–7]. How distinct these mechanisms are and, consequently, at which site in the respiratory system pathogens deposit depends on a multitude of factors. To name a few, the inhaled particle characteristics as well as the susceptible individual's anatomy and health status have an impact. It also makes a difference whether one breathes through the nose or mouth, at a high or low frequency, and with a longer or shorter pause between inhalation and exhalation. As a consequence, respiratory particle deposition could vary substantially from person to person [59], and the overall complexity prohibits an analytical description.

Therefore, the scientific community mostly relies on experimental measurements of the total deposition. This quantity sums up all fractions of deposition occurring within the respiratory tract. Regional deposition of a certain fraction of particles is disregarded, although the dose and entry point of a pathogen determine where the infection is inflamed and how severe the consequences are [59]. Fig. 2.7 visualizes a general relationship for the total deposition as a function of the particle diameter. According to this estimate, more than half of the particles with a size of 2 µm to 100 µm deposit in the airways.

The region below 0.1 µm is irrelevant because SARS-CoV-2 virions are larger. Particles between 0.1 µm and 2 µm are least likely to deposit in the airways. This is the size range mostly produced by normal breathing (see Fig. 2.4). However, this does not necessarily diminish the role of breathing in SARS-CoV-2 transmission. Particles can grow again due to condensation when entering the humid environment of human airways. Therefore, initially small particles may deposit because of an increased size [114]. However, there is a considerable knowledge gap and further research is required to reliably quantify respiratory deposition of airborne SARS-CoV-2 [59, 114].

2.2 Reactions to the coronavirus pandemic 2019

The crucial contribution of aerosols to the dynamics of the COVID-19 pandemic was not always as clear as it appears in retrospect. Several months passed before a large part of the scientific community and decision-makers gained a reasonably well-founded idea of the driving factors. In the meantime, governments and the civil population took action, with regionally differing strategies and vigor, to contain the spread while science across disciplines tried to answer urgent questions. These reactions influence the focus and the knowledge base available for the present work. Therefore, they are addressed in the following subsections.

2.2.1 Transmission mitigation and containment strategies in everyday life

In the public sector, the pandemic triggered containment and mitigation strategies. As visualized in Fig. 2.8, a major goal was to reduce the rate of new infections to prevent health care systems from being overwhelmed. An even more rigid strategy aimed at elimination. In either case, this involved far-reaching measures, which differed from region to region, were adapted over time, applied to specific subgroups of a population, and experienced varying compliance. The population sometimes acted responsibly before authorities had implemented certain measures. The contrary could also be observed where regulations were consciously or unintentionally violated. This resulted in a clutter of conditions. These circumstances have severe consequences for the recorded data and which conclusions can be drawn from them. This also means that such measures must be included in the models used to predict the spread.

Fig. 2.9 provides an overview of several measures. A prominent example is commonly known as physical or social distancing. It defines a mutual distance two individuals should keep. In this work, I prefer the term physical distancing because social distancing sometimes also refers to a set of interventions, including a reduction of social contacts one maintains. Physical distancing was imposed in many countries to reduce the transmission risk via larger droplets that could directly fall onto a susceptible person's mucus membranes. Depending on the country, the guidelines prescribed 1 m to 3 m [115]. A drawback of physical distancing is that it cannot always be maintained because people may find it difficult to estimate the distance correctly, or compliant behavior is simply impossible in overly crowded spaces [116]. Some conditions appeared especially critical due to an extended length of stay in, for example, a waiting room. To reduce the exposure

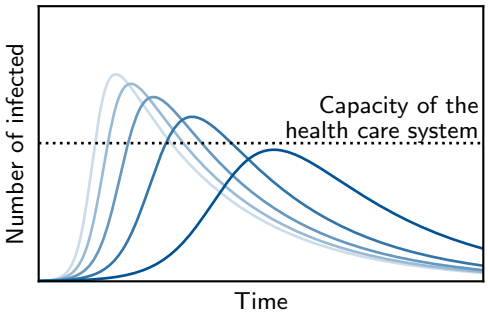


Figure 2.8: The goal of containment and mitigation measures is to flatten the curve of infected persons to a level that does not exceed the capacity of a health care system.

risk, queues were guided so that physical distancing could be maintained. Alternatively, queues were replaced by a waitlist or ticket system and fixed positions where people had to wait until they were called. Since such interventions affect human locomotion behavior, several of them are touched on in the modeling part of this work.

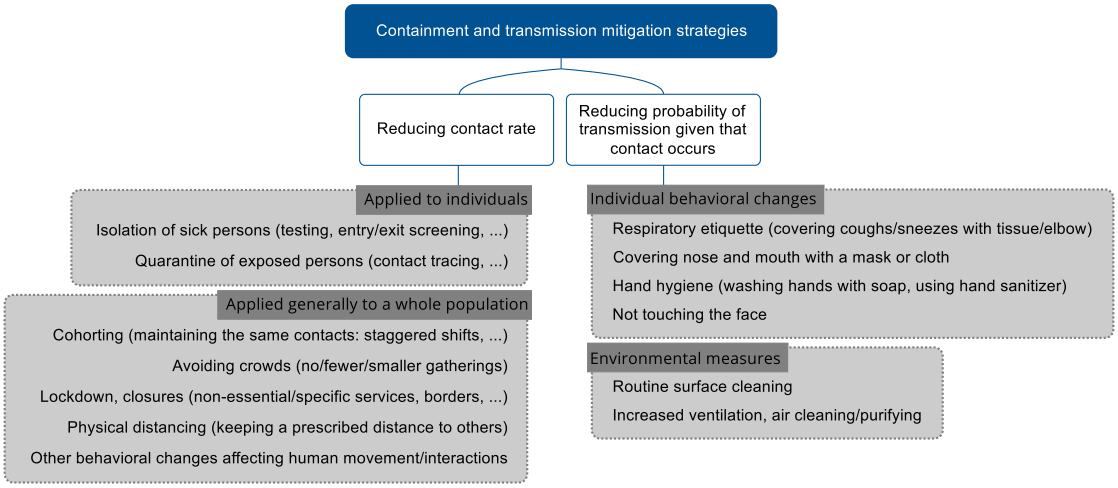


Figure 2.9: Non-exhaustive overview of non-pharmaceutical containment and transmission mitigation strategies taken during the COVID-19 pandemic. Some terms have not been defined definitely. For example, the terms social distancing and physical distancing are often used interchangeably.

For the sake of completeness, Fig. 2.9 lists additional measures, such as covering nose and mouth with a mask or contact tracing. Masks filter respiratory aerosols or droplets and, thus, can reduce the infection risk significantly [117]. Contact tracing aims to break chains of infection by identifying *close contacts* and recommending or ordering them to quarantine themselves. Close contacts are persons with a high risk of infection resulting from contact with one or more infectious individuals for a given time. The exact definition varies from country to country and was updated during the course of the pandemic. The

Robert Koch-Institut [20], Germany's central scientific institution for safeguarding public health in the context of the pandemic, devised the following definition:

Definition: close contact

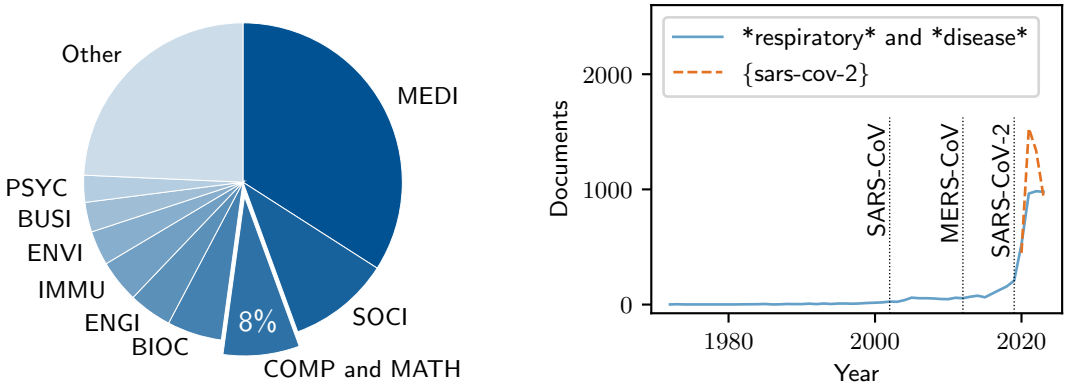
Close contacts are persons who stay in close proximity, less than 1.5 m, to a confirmed COVID-19 case for over 10 min without adequate protection through masks.

Other conditions can also suffice to reach the same exposure risk. Digital contact tracing frameworks, such as the German Corona-Warn-App [118], typically evaluate the exposure risk based on exposure time and proximity. The effectiveness of wearing masks [117], contact tracing [119, 120], and other measures [121] has been reviewed manifold. They are not further investigated in this work. However, their number and impact show how seriously daily life was restricted throughout the COVID-19 pandemic. This points out the importance of making measures bearable and not unnecessarily restrictive. To support informed decisions and tailored actions against the spread, the scientific community tried to answer open questions.

2.2.2 Impact of the pandemic on scientific research

The COVID-19 pandemic called forth an unprecedented response by the scientific workforce to various aspects of the same overarching problem. The magnitude of this reaction manifests itself in the sheer quantity of contributions originating from various fields of research and published within a short period of time. As a result of a document search on Scopus that I conducted (see Appendix B for details), Fig. 2.10 provides a simplified overview of the multidisciplinary interest from many subject areas and an increasing number of contributions from computer science and mathematics related to modeling and simulation of respiratory diseases. Of course, these figures do not reflect the full picture because the search results depend on the database. For example, Scopus ranks among the largest databases, but it does not index all COVID-19 literature [122]. Furthermore, search terms can change over time, as is the case for SARS-CoV-2, which was temporarily named novel coronavirus (2019-nCoV) [28]. Therefore, my search may include irrelevant or exclude relevant contributions, but the figures illustrate three major challenges influencing this work:

- The vast number of scientific publications makes it impossible to detect and include all relevant publications on recent developments. Fig. 2.10b shows slightly higher numbers of annual publications for the years after 2000. This increase in attention was possibly promoted by outbreaks of respiratory diseases such as SARS or MERS, influenza waves, and others. The drastic inclination after 2019 is clearly associated with the outbreak of SARS-CoV-2. The decline since end of 2021 continues until 2023, but the annual output is still higher than before the outbreak.
- The pressure to find solutions to burning questions was high, so research questions were quickly addressed and findings were published early so that others could benefit. A story of success for accelerated generation and acquisition of knowledge is



(a) Number of documents by subject area; search term: *{sars-cov-2} or {covid-19}*; 567 023 documents listed in total; acronyms are listed in Table B.1.

(b) Number of documents published annually in the subject areas computer science and mathematics; search terms: **model* or *simulation** combined with **respiratory* and *disease** and *{sars-cov-2}*; 4911 and 4257 documents listed in total, respectively.

Figure 2.10: The diagrams depict the number of documents indexed on Scopus with published dates before 2024 and matching search terms in article title, abstract, and keywords. I conducted the search on January 2, 2024.

the rapid development of effective COVID-19 vaccines [123]. Such tendencies of compressed research activities also become apparent, for example, in the shape of increasingly cited material that has been published before passing peer review, so-called preprint content [124]. However, this also raises concerns about the quality of research [125].

- An interdisciplinary subject affords an understanding of multiple fields of research. Most of them offer various approaches to answer similar questions in a different manner. The difficulty is selecting suitable approaches and pairing them. In the context of COVID-19, interdisciplinary work also implies bridging knowledge gaps and establishing or renewing shared vocabulary, for example, for transmission via respiratory aerosols.

My strategy to handle these challenges is, first of all, to acknowledge that, although COVID-19 literature is technically accessible, an encompassing review of state-of-the-art and recently gained knowledge exceeds the scope of this dissertation by far. Focusing on publications that address similar research questions or treat them in a comparable manner produces some relief, but the remaining publication base is still unmanageable. Therefore, I prefer review studies over primary sources, at least for the medical and health-related subtopics. Review studies synthesize several studies at once and can balance out the imprecision of a single study. They also contextualize outdated findings, for which reason I conducted a repetitive literature search and attached higher value to the

latest reviews if available. Some reviews contrast findings from different fields of research and, thus, provide insight into aspects from relevant disciplines. Finally, I limit the considered subtopics and tools to selected fields. These are computer science, mathematics, physics, medicine and health, and their established interdisciplinary combinations, such as scientific computing.

2.3 Reported data on superspreading events

Now that the consequences of the COVID-19 pandemic for everyday life and the scientific landscape have been addressed, it is easier to appraise scientific reports and evaluate where they possibly lack information or contain outdated statements. In the following, I examine documented data on superspreading events because these events have been identified as a driving factor for the spread of COVID-19. Moreover, in the absence of other data, superspreading events are crucial for developing and validating a model.

As introduced in Section 2.1.2, superspreading events are characterized by a sudden increase of infections [5, 126]. At least one superspreader is responsible for an extraordinarily high secondary attack rate [126]. Typically, the number of secondary infections ranges between 10 to 10^2 [127, 128], but there is no definite minimum number of secondary infections.

Superspreading events have been recorded throughout the pandemic. Majra et al. [75] provide a non-exhaustive collection of such infection clusters. In their overview, typical settings of spreading events are worker dormitories, workplaces, schools, hospitals, elderly care, religious gatherings, and other leisure activities, such as shopping or going to a bar. I explicate two spreading events among people going to a restaurant and attending a choir practice. These scenarios provide a basis for the modeling part of this dissertation. They are particularly interesting because the events occurred in January and March 2020, respectively. At this time, little was known about SARS-CoV-2, and measures did not distort the typical behavior and socializing.

The spreading event in a restaurant in Guangzhou, China, was first analyzed by Lu et al. [73] and, based on their report, reconsidered multiple times. One infectious person caused nine further infections in three families in a restaurant. The three families stayed in the restaurant approximately at the same time at three tables next to each other. The index case's family shared the space with the other two families for an overlapping time of 53 min and 73 min, respectively. During this time, the index case infected at least one person from the two other families. It cannot be excluded that the remaining members of each family were infected outside of the restaurant.

The other spreading event took place during a choir rehearsal in Skagit County, Washington [71, 129]. One among 61 attendees was symptomatic, who was later tested positive along with 32 other participants. Additional 20 persons were declared probable cases without being tested. However, some of these could erroneously be identified as secondary cases, as it happened for one person who was first considered infected with SARS-CoV-2 but tested negative after symptom onset. Overall, 32 to 52 participants of the attendees were infected, which results in a secondary attack rate of 53.3 % to 86.7 %.

The choir rehearsal took place in two rooms with varying seating orders. However, the exact arrangement of the seats and the participants' positions are unknown. They stayed all in the same room with some seats remaining unoccupied for the first 45 min. Then, the choir was divided, and each group practiced in a separate room for another 45 min. The infectious person was part of the group that stayed in the room where the practice started. The sectional rehearsal was followed by a break of roughly 10 min. During the final 50 min, the practice was conducted in the same setting as the initial session.

Overall, data on superspreading events is available, but the more recently the spreading event occurred, the more behavioral changes, immunization, and virus variants come into play. In turn, the earlier the spreading event occurred, the less aware the population and scientific community were. This could have resulted in recording or providing fewer details, compromising the trust in the data.

2.4 Summary

Human-to-human transmission of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) occurs mainly through respiratory aerosols. Aerosol particles cover diameters from 0.002 µm to over 100 µm of which the particles above a size of approximately 0.1 µm can carry viable SARS-CoV-2. The particle size determines the behavior of an aerosol. Larger aerosol particles typically originate from violent respiratory events, for instance, coughing. Because of their size, they settle faster than they evaporate. Consequently, transmission via larger aerosol particles is bound to shorter ranges or fomites. Smaller aerosol particles are produced by any respiratory activity, including breathing. They remain airborne for several minutes to hours and can spread spatially over longer distances. Hence, evaporation is a dominant effect, making the size of aerosol particles highly dynamic. Several researchers argue that the role of smaller aerosol particles is more critical, which suggests that breathing belongs to the primary transmission mechanisms. However, the exact relative contribution of small and large particles to the dynamics of the disease spread has not been analyzed conclusively. Furthermore, many questions about aerosol generation, the persistence of exhaled aerosols, and their deposition in the lungs remain unanswered so far, despite collective efforts of the scientific community.

The increased interest across disciplines in SARS-CoV-2 transmission resulted in an extraordinarily high quantity of related academic publications. It also temporarily accelerated the publication process, which may negatively affect the quality. Moreover, fundamental findings, such as the notion of the dominant transmission path, have changed since the emergence of the virus. To cope with these issues, I referred to review studies where available, treated the literature carefully, and acknowledged uncertainties or limitations. Importantly, large parts of empirical data about SARS-CoV-2 transmission are biased because containment strategies implemented during the pandemic distorted the records. Reports of superspreading events often omit relevant information. Laboratory experiments are usually out of the question for ethical reasons. This strengthens the need for modeling and simulation as a basis for decision-making.

3 Mathematical modeling of infectious diseases

Estimating the spread of infectious diseases with the aid of computer simulations requires a suitable model. This involves finding mathematical formulations of the transmission mechanisms described in the previous chapter. This chapter presents the state of the art and recent developments on how these mechanisms are operationalized and simulated. Infectious disease modeling refers to both inter- and intra-individual spread of infection. Models that describe transmission from one person to another belong to the category of epidemiological models. In contrast, models considering an individual's immune response or pathogen load are often called within-host models [130, p. 5347]. I provide an overview of epidemiological models operating on distinct or multiple scales, some combined with within-host approaches.

Epidemiology is a branch of medical science that deals with the causes, spread, and consequences of diseases affecting the health of a whole population. Accordingly, epidemiological models provide mathematical descriptions of how an infectious disease outbreak evolves in a population. These mathematical formulations and the resulting predictions improve the understanding of the dynamics of spreading diseases and support epidemiologists in identifying effective strategies for infection control. Such models were heavily used during the COVID-19 pandemic (see Fig. 2.10b) and during previous disease outbreaks. As with any model, they are built upon simplifying assumptions and are subject to uncertainties, which delimits their predictive power.

The following sections are structured according to the scale of the epidemiological models. Among other possible classifications, my definition of large-scale and small-scale infectious disease models is an attempt to emphasize the setting of the problem and the purpose of a simulation study. However, the multitude of approaches developed during the COVID-19 pandemic yields a gradually shifting scope.

I approach the topic top-down, starting with the large-scale perspective. Here, large-scale means that a model treats populations in a relatively coarse manner, for example, divided into sub-populations with shared attributes. The observed population is usually rather large, ranging from global or national to regional levels, and the predictions cover days to years. The quantities of interest, such as an infection rate, are macroscopic because they are valid for a whole population or sub-population. Thus, one often pursues the questions: Which sub-groups contribute significantly to the spread, and where are epidemic hot spots?

The subsequent considerations zoom into the details of between-host transmission to finally arrive at finer scales. A finer resolution means that the simulation is restricted to small populations of two up to a few hundred individuals. This restriction often occurs because of increasing complexity and computational expense. Proportionately, the spatial extent of the considered scenarios and the simulated time shrink to the size of rooms or buildings and a few minutes to hours, respectively. This level of granularity

allows for reenacting and scrutinizing superspreading events, a driving factor of the spread of COVID-19. Questions stated in this context direct toward: How does inter-individual transmission work? The mission is to harness this knowledge to mitigate individual infection risks. Since many small-scale models, including the one I introduce in the second part of this dissertation, integrate crowd models, I briefly recall the concept of microscopic crowd simulation. Finally, I summarize the essential advantages and disadvantages of each modeling approach.

3.1 Large-scale models

Epidemiological models covered by this section operate on a larger scale and provide macroscopic information on the spread of an infection within a larger population, for example, the inhabitants of a region or country struck by an epidemic. This model type typically aims to predict the proportion of infected individuals or the duration of an epidemic. It allows us to derive epidemiological key figures. The basic reproduction number is a well-known example. First mathematical descriptions evolved in the early 20^{th} century and provide the basis for further developments, for example, accounting for sub-populations, specific social interactions, and mobility patterns. In this section, I introduce these models, starting with a simple deterministic form and successively increasing the complexity.

3.1.1 Compartmental models

Compartmental models represent a standard method for tracking the proportion of susceptible, infectious, and recovered members of a population over time. Their basic concept dates back to the ideas of Hamer [131], Ross [132], Ross and Hudson [133], and Kermack and McKendrick [134]. In their cornerstone work, Kermack and McKendrick [134] formulate a deterministic relationship between the compartments of susceptible, x, infectious, y, and recovered, z, people in a closed, homogeneously mixed population ($N = x + y + z = \text{const.}$) over time, t. They assume a constant transmission rate, κ, and recovery or death rate, l, per time unit to link the compartments through the following differential equations:[1]

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\kappa x y \tag{3.1}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \kappa x y - l y \tag{3.2}$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = l y \tag{3.3}$$

These basic equations and some of their extensions can be solved analytically [135]. Numerical approaches using various methods, for example, provided by solvers for ordinary differential equations or partial differential equations, are common [136].

[1] Another common notation, but originally not used in [134], replaces the transmission rate by the rate of new infections an infectious person causes, $\beta = \kappa N$.

With $\frac{\mathrm{d}y}{\mathrm{d}t} > 0$, we obtain an essential epidemiological parameter:

$$R_0 = \frac{\kappa N}{l}. \tag{3.4}$$

It is termed the basic reproduction number and describes the average number of secondary cases infected by an index patient in a completely susceptible population ($x \approx N$). If $R_0 > 1$, an epidemic will occur, as visualized in Fig. 3.1. The effective reproductive number applies to a partly immune population. It is defined as

$$R_e = R_0 \frac{x}{N} = \frac{\kappa x}{l}. \tag{3.5}$$

The reproduction number is employed, often in an adapted version for specific conditions, to compare the infection risk in certain scenarios or to evaluate the effectiveness of measures against the spread of COVID-19.

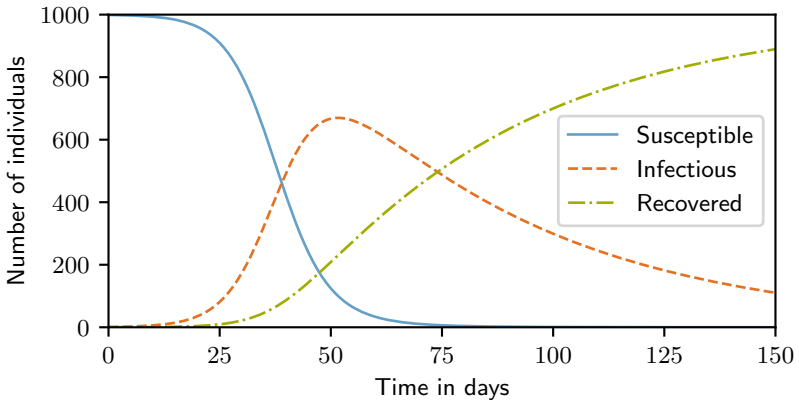
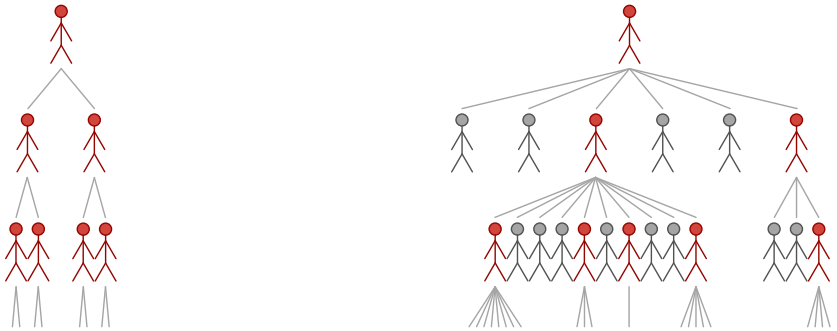


Figure 3.1: The SIR model (Eq. 3.1-3.3) yields the number of susceptible ($x(0) = 999$), infectious ($y(0) = 1$), and recovered ($z(0) = 0$) persons for $\kappa = 0.0002\,\mathrm{d}^{-1}$ and $l = 0.02\,\mathrm{d}^{-1}$.

Despite the informative value of the reproduction number, it should be noted that, in its general form, it represents an average quantity. Therefore, it has limited validity when the infection spread is examined for a small population at the level of individuals. Fig. 3.2 compares two chains of infection in which every member of the population (red and gray) infects two other persons on average. The average value is an appropriate measure for the condition in Fig. 3.2a, but it can be an oversimplification of the conditions in Fig. 3.2b, where only 8 individuals (20 %) are responsible for 40 infections (80 %).

The classic susceptible-infected-removed (SIR) model can be extended with additional compartments and connections between them as they best characterize the spreading disease's properties and serve the model's purpose [137]. There is practically no limit to introducing further compartments and sub-populations. For example, the latent period of a disease is typically represented by an exposed state [137]. The population is divided into age groups to capture age-dependent characteristics of a spreading disease [137].



(a) No variability in the number of new infections caused by individuals.

(b) Significant variability the in number of new infections caused by individuals.

Figure 3.2: The problem with the reproductive number: The same reproductive number describes the two scenarios of a disease outbreak differently well. All stick figures represent infected individuals among which only the red ones infect others.

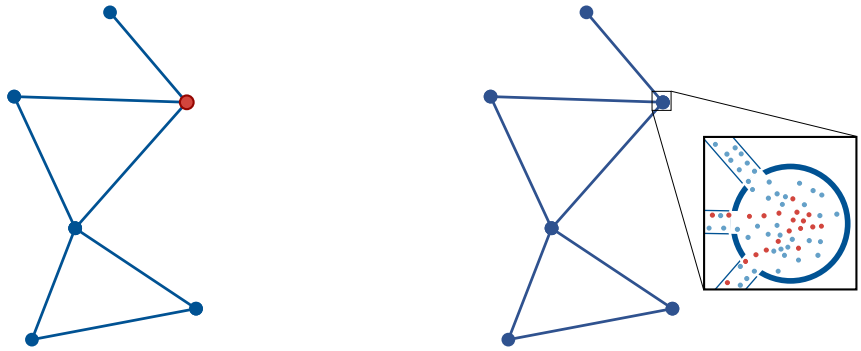
A spatial component and specific contact patterns can be used to discard the strong assumption of homogeneous mixing within the population. Such refinements are steps toward network or other individual-based approaches. For a detailed mapping of the SIR model to complex networks, I refer to [138].

3.1.2 Network models

Network models in epidemiology are often used to determine macroscopic quantities such as infection rates in certain sub-populations or specific regions. To obtain such information, a network represents individuals or entities as nodes and their connections as edges (see Fig. 3.3).

If each node corresponds to a single individual (Fig. 3.3a), the first task is to decide which nodes should be linked. Retrieving data for an actual population, for example, through contact tracing, is complicated and can raise data protection issues. Once the nodes have been connected, one must define how these connections represent transmission. In the simplest case, an edge is regarded as existent or non-existent. A more complex method introduces weights depending on the strength of the connection [139]. In simulations of airborne disease transmission, the edges embody social interactions. These contact patterns are temporally heterogeneous, which can be analyzed with temporal networks [140]. Such epidemiological network models simulating the spread of airborne diseases are presented, for example, for smallpox [141] or for COVID-19 outbreaks on a cruise ship [142] and at workplaces [143].

An alternative approach accounts for mobility patterns and, for this purpose, defines nodes as locations that can be occupied by multiple individuals at the same time (see Fig. 3.3b). Edges connecting two locations allow individuals to commute between them [140]. Already before the outbreak of COVID-19, many network models, such as the ones published in [144, 145], have been developed to unravel the spatial component of disease



(a) Individuals (nodes) connected through social interactions (edges).

(b) Entities (nodes) connected through paths (edges) along which individuals (small dots) can reach other entities.

Figure 3.3: The schematically illustrated network models define nodes as (a) individuals or (b) entities. The latter can contain more than one individual. Depending on the scale of the model, these entities stand for, for example, rooms, whole buildings, or larger units. In both models, red dots indicate infectious individuals.

transmission in an increasingly globalized world. More recent contributions addressing the spread of COVID-19 on different scales can be found, for instance, in [146–149]. While Chinazzi et al. [146] use a global model that spans approximately 200 countries and territories, Chang et al. [147] and Kühn et al. [148] analyze the spread on a national level. They connect multiple SIR-like models for sub-populations of a country through a graph. The simulations in [147] rely on mobility data for metropolitan areas in the United States. Kühn et al. [148] use their high-performance software MEmilio [150] to predict the course of the COVID-19 pandemic in Germany, taking into account demographic data. The simulations confirm superspreading events as driving factors. However, the models are not capable of resolving heterogeneity in the transmission below the level of a sub-population where spreading events occur [147, 148]. In a similar manner, Müller et al. [149] collate real mobility data and census data to schedule activities of a synthetic population within the metropolitan area of Berlin, Germany. Their work is based on large-scale traffic simulations with the Multi-Agent Transport Simulation (MATSim) tool [151]. The epidemiological feature of the simulation framework is defined by an exponential dose-response model that describes an individual's infection probability, given that contact with an infectious person is established at that node. With this type of model, one typically predicts the number of persons that become infected, hospitalized, and require intensive care. This provides a basis for calculating, for example, the reproduction number per sub-group or activity type.

The above models have some limitations. Above all, calibrating them is difficult because they depend on a multitude of input parameters. Thus, considerable uncertainties percolate through the simulation. These uncertainties need to be treated adequately, for example, with Monte Carlo simulations to obtain reliable results. This quickly becomes

computationally expensive for large populations. The literature often does not report the computational resources required. For example, Müller et al. [149] only state that their model is computationally demanding. This could be the reason for a relatively small sample size of eight repeated and averaged simulation runs. Furthermore, quantities such as the reproduction number per sub-group or activity type represent average numbers and do not reveal information about the infection process for a specific situation. On the level of unique everyday situations, small-scale approaches come into play to solve this problem.

3.2 Small-scale models

This section addresses models for between-host disease transmission, acting on local scales. In this context, the population is fairly small, typically with numbers in the order of magnitude of persons working in an office, using a specific vehicle of public transportation services, going to a restaurant, or any other store. Because of the small population size, it is possible to attribute distinct properties to each individual. The model output and the drawn conclusions are usually valid for the one scenario that has been simulated. Transferring them to similar scenarios can be improper. A common purpose of between-host models is to better understand the physical transport phenomenon of pathogens passing from an infectious to a susceptible person. Based on that, one can calculate each individual's exposure or infection risk, the total exposure time, the number of exposed and infected individuals, or a reproductive number for that scenario. These quantities facilitate evaluations and comparisons of how critical the simulated situations are.

The following subsections distinguish between approaches that assume homogeneous and inhomogeneous exposure. In contrast to the large-scale perspective, the term (in)homogeneity now does not refer to (in)homogeneous mingling of individuals. Instead, (in)homogeneity determines whether the pathogen is assumed to be uniformly distributed across the considered space. This section closes with a brief excursus about microscopic crowd models because they constitute the foundation of several recently developed disease transmission models, including my work.

3.2.1 Homogeneous exposure risk

Probably the most prominent model type that assumes homogeneous exposure to airborne contaminants in closed environments is based on the Wells-Riley equation [6, 7]. This model can be approached as follows (for example, see [152]): We revisit the idea of the compartmental SIR model and consider I infectious persons entering a room with S susceptible persons. The infectious persons release airborne pathogens, which instantaneously and evenly disperse across the room. After time t, all individuals leave the room. Rewriting Eq. 3.1 and adding the number of persons exposed to airborne pathogens, E,

with an incubation rate, α, yields

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\kappa(t)SI \tag{3.6}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \kappa(t)SI - \alpha E. \tag{3.7}$$

The change in the number of infected and recovered persons can be omitted because the considered time interval, τ, is much shorter than it takes an exposed person to become infectious ($\alpha\tau \ll 1$) or an infectious person to recover. Therefore, the term αE is negligible, and the number of infectious persons, I, remains constant. The transmission rate, κ, is now time-dependent since the pathogen concentration in a well-mixed room changes when the infectious person enters. Substituting $\kappa(t)$ with $\mathrm{d}\hat{t} = \kappa(t)\mathrm{d}t$ and using the conditions $S(0) = N - I$ and $N = S(t) + E(t) + I$ delivers the solutions

$$S(t) = (N - I)e^{-q(t)} \tag{3.8}$$

$$E(t) = (N - I)(1 - e^{-q(t)}), \tag{3.9}$$

where $q(t) = I\hat{t} = I\int_0^t \kappa \mathrm{d}t$ denotes the infection quanta per time produced by the infectious individuals.

Originally, Wells [6, pp. 140–4] proposed the concept of quantum of infection. According to his definition, a quantum equals the amount of pathogen that leads to a chance of infection in $1 - e^{-1} = 63.2\,\%$ of the susceptibles [6, p. 140]. Based on these considerations, Riley et al. [7] formulate the probability of infection by

$$P = 1 - e^{-Ir} \text{ with } r = \frac{qpt}{Q}. \tag{3.10}$$

This equation takes into account the rate of quanta released per infector, q, the volume of air ventilated by the susceptible persons, p, and steady-state room ventilation at rate Q. The sum of the individual probabilities of infection for all susceptible persons resulting from independent exposure in n environments returns the number of new cases

$$C = \sum_{i=1}^{S} 1 - e^{-(I_1 r_1 + I_2 r_2 + \ldots + I_n r_n)}, \tag{3.11}$$

known as the Wells-Riley equation, which is equivalent to Eq. 3.9 for $n = 1$.[2]

It should be noted that Wells-Riley-like models implicitly evaluate both inter-individual transmission and pathogen-host interaction. As explained in [62, pp. 269–71], this can be separated into the processes of exposure and dose-response. More precisely, the exponential dose-response relationship formulated in Eq. 3.10 is composed of the probability of ingesting a certain amount of pathogens, P_1, and the probability that the ingested pathogens cause an infection, P_2. That is, an individual ingests j pathogens from an

[2] In the literature, one often finds the following version of the Wells-Riley equation: $C = S(1-e^{-Iqpt/AV})$, with A denoting the air exchanges per hour and V indicating the room volume [153].

exposure in volume V with mean pathogen density $\bar{\mu}$. Hence, the expected average dose d equals $\bar{\mu}V$. For randomly distributed pathogens at low densities, the probability of ingestion is characterized by the Poisson distribution:

$$P_1(j \mid d) = \frac{d^j}{j!}e^{-d} \tag{3.12}$$

The subsequent process of k ingested pathogens surviving to initiate infection with independent and identical probability λ is often defined by the binomial distribution:

$$P_2(k \mid j) = \frac{j!}{k!(j-k!)}(1-\lambda)^{j-k}\lambda^k \tag{3.13}$$

An alternative, biologically plausible dose-response model represents P_2 by the beta distribution or its approximations [62, pp. 272–4]. Depending on the pathogen and host susceptibility, this can but does not necessarily provide better estimates than the exponential dose-response model [10]. Given that infection occurs when at least $k_{min} = 1$ pathogens survive, the joint probability of P_1 and P_2 returns the probability of infection:[3]

$$P_I(d) = \sum_{k=k_{min}}^{\infty}\sum_{j=k}^{\infty} P_1(j \mid d)P_2(k \mid j) = 1 - e^{-\lambda d}. \tag{3.14}$$

This formula is equivalent to Eq. 3.10, but other than the Wells-Riley approach, this generalized denotation is not limited to airborne pathogen transmission. More complex dose-response models that consider, for example, variable host sensitivity are presented in [62].

The Wells-Riley model has been refined in various ways to soften some strong assumptions that it entails. For more details, I refer to the reviews [9, 153], which examine further developments of the Wells-Riley model. Among these improvements are, for example, changing quanta levels over time [154], terms incorporating respiratory protection [9, 155, 156], and declining pathogen concentrations through air cleaning [156], aerosol deposition, or shrinking pathogen viability [9].

In the context of the COVID-19 pandemic, these improved models were heavily used. They provided the basis for deriving quanta emission rates or estimating the infection risk associated with certain situations, for example, a superspreading event during a choir rehearsal [71]. Another study analyzes generalized scenarios of several customers entering different types of indoor environments, such as retail stores or service providers [157]. The scenario comparison in [157] emphasizes that proper ventilation combined with reduced occupancy levels achieved by obliging customers to queue outside the store can significantly reduce the infection risk. Note that these publications and typically also other studies that use the Wells-Riley approach focus on mathematical modeling and investigating the model output. The underlying implementation or details about the simulation set-up are often not discussed.

[3] k_{min} is not equivalent to the minimal infectious dose. If $k_{min} > 1$, one obtains a threshold model [62, pp. 270–6], which relies on the biologically implausible assumption that pathogens take effect collectively as a group [9].

Despite improvements of the exponential Wells-Riley equation, several limitations still need to be addressed. Fully validated dose-response models for human-to-human transmission of SARS-CoV-2 are currently not available. One reason is the lack of experimental data [10]. When it comes to airborne transmission of respiratory diseases in general, the knowledge gap related to respiratory deposition is responsible for the constrained predictive power of the model. As explained in Chapter 2, respiratory deposition of aerosol particles represents a considerable source of uncertainty. Nevertheless, the deposition is not explicitly incorporated in the model [59]. Additionally, the Wells-Riley equation ignores inter-individual variance in host susceptibility.

Finally, an evident and recognized limitation of the Wells-Riley model is the assumption of an instantaneously well-mixed room. This assumption is grounded in the observation that airflows, driven naturally or by mechanical ventilation, act particularly on smaller aerosol particles. These flows are turbulent, and turbulence tends to accelerate the homogenization of airborne particles in indoor spaces [158]. The well-mixed theory implies that the degree of exposure is equal for all individuals. Depending on the airflow conditions and room geometry, this can constitute a major simplification [159]. Salinas et al. [158] show that the well-mixed condition can underestimate a susceptible person's exposure for distances smaller than 5 m to an infector and vice versa. Presumably, no reasonable ventilation rate would produce entirely mixed air [158]. Indoor settings with significant spatiotemporal differences in the pathogen concentration exist beyond doubt. Otherwise, several COVID-19 infection clusters would seem unlikely [70]. To address this problem of inhomogeneously distributed pathogen concentrations, Martinez et al. [160] consider rooms with distinct quanta levels and introduce virtual persons with individual schedules of activities in different rooms. Other recent developments focusing on airborne transmission of SARS-CoV-2 depart from the well-mixed condition and resolve spatial variations, for example, by means of rules determining pathogen levels within a grid of homogeneously mixed cells [161], a semi-analytic description [159], or a Lagrangian and Eulerian specification of the flow field [162].

In summary, many models estimate infection rates for specific daily-life situations based on the Wells-Riley equation. The Wells-Riley equation assumes that airborne pathogen loads are distributed homogeneously in a room, and it estimates the probability of infection as a negative exponential function of the inhaled dose. However, in the case of SARS-CoV-2 transmission, there is a considerable lack of data and plausible dose-response models, which complicates or even prohibits calibrating such a model. Furthermore, Wells-Riley-like models can be a reasonable simplification for some scenarios, typically for airborne transmission in ventilated indoor environments. Nevertheless, they do not provide insight into the mingling or interaction between individuals and the respective transmission characteristics. Such a level of detail requires that the homogeneous assumption is discarded and spatial variations in pathogen concentrations are taken into account.

3.2.2 Inhomogeneous exposure risk

A growing body of literature recognizes inhomogeneous transmission characteristics in crowds of moving people. Since the transmission mechanism is resolved spatially in these models, it also makes sense to take a closer look at the virtual persons' positions. Many daily-life situations are rather dynamic in the sense that people move around. Therefore, this model type often considers both pathogen transmission and human movement. This section focuses on the former, that is, how transmission under the inhomogeneous assumption is cast into a mathematical description. For a few reasons, the section is much more extensive than other passages in this state-of-the-art chapter. Firstly, none of the subsequently examined approaches has been fully established yet. Secondly, no far-reaching review is available on this subject matter. And thirdly, my own work belongs to this model category, as will become apparent in Part II. Therefore, I investigated how inter-host transmission has previously been modeled under the inhomogeneous assumption and what the respective strengths and weaknesses of such methods are. This involved identifying key variables related to disease transmission models, proposing a suitable classification, and discussing the underlying concepts and several seminal works.

3.2.2.1 Review method

I conducted a systematic and a targeted literature review. Following the method in [163], I aimed at aggregating important literature and reducing the likelihood of bias. An exhaustive systematic review exceeds the scope of this dissertation because of the large database of COVID-19 literature. Consequently, as with any review, my search process may have failed to identify relevant articles. To iron out gaps in the systematic approach, I decided to include the results of a targeted review.

The analysis consisted of collecting, extracting, and categorizing bibliographic data retrieved from Scopus. I chose Scopus because it belongs to one of the largest abstract databases for publications, meeting the requirement for peer review quality. The search encompassed the following steps (see Appendix C for more details):

(1) Searching relevant literature, for example, following a pyramid scheme to gain an understanding of the subject and to identify terms commonly used;

(2) Synthesizing frequently used keywords and defining search strings; Deciding on exclusion criteria for screening the search results;

(3) Conducting the document search on Scopus;

(4) Collating search results found through the targeted review and step (3);

(5) Removing duplicates; Applying inclusion and exclusion criteria within a two-step screening process (title and abstract screening followed by a full-text review);

(6) Data extraction, appraisal, and reporting;

Note that the targeted review in step (4) essentially comprised checking the reference lists of documents obtained in step (3) or collecting publications found through step (1). Admittedly, such a procedure can distort the results, but a bias appears less problematic here since the data appraisal focused on the general modeling approach presented in a publication and not on quantitative data. Most publications found in this manner escaped the systematic search because of the same criterion. They did not fit the search string that limits the results to literature associated with the relevant transmission paths (see number S4 in Table C.1). In a different context, such a constraint of a systematic literature review has been admitted, for example, by Templeton et al. [164] although experienced scientists were involved in defining the search terms.

The screening process included 125 publications, 124 without duplicates, retrieved from Scopus, of which 32 met all inclusion requirements. The targeted review delivered 20 additional contributions. The final stage of the synthesis was to categorize the literature according to the underlying models for disease transmission.

3.2.2.2 Overview of common modeling concepts

I identified four categories of how disease transmission is operationalized. Exposure or infection risk is evaluated based on

a) the proximity to infectious individuals (see Fig. 3.4a),

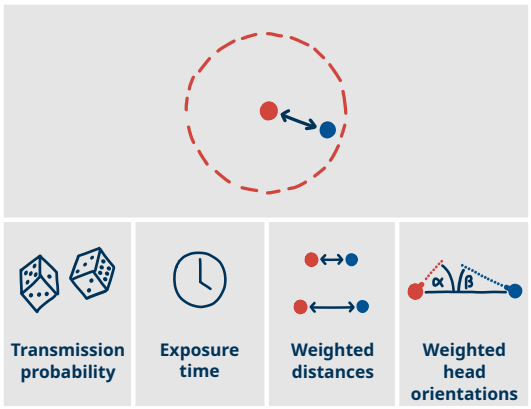
b) a medium of pathogen transmission such as aerosols (see Fig. 3.4b),

c) a combination of the proximity and transmission medium approach, and

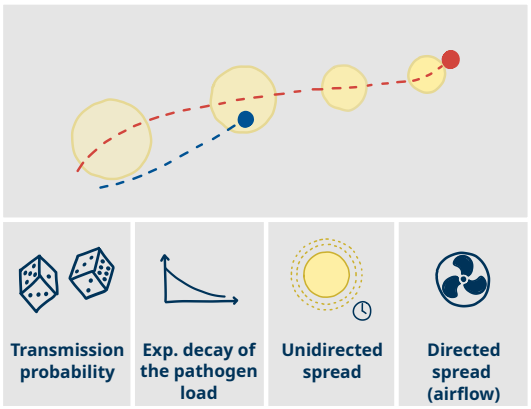
d) other means that belong to none of these categories.

The majority of the analyzed publications ($n = 30$, 57.7 %), particularly those issued before the outbreak of COVID-19, adopt the proximity-based approach. That is, the transmission is expressed as a function of the mutual distance, d, between infectious and susceptible individuals. Typically, a cut-off distance of $d_e = 1\,\mathrm{m}$ to $2\,\mathrm{m}$, in a few cases even $3.5\,\mathrm{m}$ or more, determines the minimum proximity required for an increase in exposure, sometimes in combination with a threshold for the minimum exposure time, t_e, this distance is maintained. This concept of proximity is equivalent to the one well-known from mobile applications developed for digital contact tracing during the COVID-19 pandemic, for example, the German Corona-Warn-App [118]. Most of these models predominantly address short-range transmission via larger respiratory particles. However, not all authors explicitly state this limited scope in their works.

Models classified by exposure to transmission media such as aerosols or fomites ($n = 10$, 19.2 %) evaluate the susceptible persons' exposure to the prevailing contamination around their current position. The contamination originates from infectious individuals releasing pathogens in a continuous or discrete manner while moving through a virtual space. Such operationalizations generally attempt to describe transmission via fomites or smaller respiratory particles. Some models consider the spatial extent of pathogen concentrations



(a) Exposure via proximity with optional features



(b) Exposure via transmission media, for example, aerosols or surfaces, with optional features

Figure 3.4: My classification of small-scale models for assessing inhomogeneous exposure risks in indoor spaces: The models are typically based on the proximity between infectious (red) and susceptible (blue) individuals (a). Alternatively, the models interpose a medium of transmission between the emitter and receiver (b). These two basic concepts are sometimes complemented with additional features, such as a transmission probability or an exposure time.

to be constant for fomite and airborne transmission, whereas others account for aerosol propagation in unventilated or ventilated environments.

A few publications ($n = 7$, 13.5 %) combine the concepts of proximity and transmission media. The remaining contributions ($n = 5$, 9.6 %) could not be attributed definitely to one of the above categories. The complete list of reviewed scientific contributions and their respective classifications, descriptions of the analyzed scenarios, and the quantities of interest are presented in Table C.2. I now discuss several of the listed works that exemplify the above categories and, in my view, have promoted the development of this model type through their novel ideas.

3.2.2.3 Examples of modeling exposure via proximity

One of the first extensions of a microscopic crowd model by a disease transmission model was proposed by Johansson and Goscé [11] in 2012. The authors simulate the spread of a fictitious disease in a dense crowd, assuming that the disease has a certain probability of jumping from an infectious to a susceptible person if the two individuals are in physical contact. Infected persons become immediately infectious themselves. Such a setting is unrealistic mainly because it neglects the latent period. Nonetheless, this contribution has put the coupling of crowd dynamics and epidemiological modeling up for discussion, opening up a new perspective on preventing the spread of communicable diseases at mass events.

A few years later, Namilae et al. [167] added a temporal component to the proximity-based approach. They specify the number of inter-individual contacts to identify preferential airplane boarding strategies in light of potential disease transmission. A contact is recognized when a minimum distance is maintained for a particular time. In addition, smaller and larger contact distances are associated with distinct modes of transmission. Nonetheless, the model remains relatively coarse and does not account for larger droplets following a certain direction after exhalation.

More recently, Garcia et al. [180] refined the proximity-based approach for short-range transmission. They derive the transmission rate from computational and experimental studies as a function of the time an infectious and a susceptible person spend together, their distance, and their respective head orientations. Applying this model to pedestrian tracking data allows the authors to compare the predicted infection rates for different daily-life situations. The transmission function is enhanced in [191] such that ambient airflows can be incorporated into the risk assessment. The results of computational fluid dynamics (CFD) simulations for various flow conditions are transferred to coarse maps of pathogen concentrations around an infector. In this manner, the model accounts for transmission risks weighted by radial and angular coordinates, but it still belongs to the proximity-based approaches because the maps are linked to a pedestrian's position. Nicolas and Mendez [209] demonstrate how to couple the maps with crowd models. The transmission maps are publicly available and can, in principle, be transferred to other scenarios. Unfortunately, the model is yet in a prototypical stage and does not capture long-range transmission.

The general concept of proximity-based approaches had already been established by the time of the outbreak of the COVID-19 pandemic. However, the transmission characteristics of SARS-CoV-2 made clear that including solely the distance and contact time between an infectious and a susceptible individual may not be enough for all diseases. An infectious person can exhale pathogens carried by respiratory particles that remain airborne for extended periods. Therefore, others entering a contaminated area potentially become infected even if the infectious person has already left the considered space. Nonetheless, a large fraction of models developed recently after the outbreak maintain the concept of proximity.

3.2.2.4 Examples of modeling exposure via a transmission medium

Evaluating exposure to infectious particles lingering in the air can be achieved by introducing a transmission medium, such as aerosols. Many models ignore the impact of ventilation and air currents on the transmission medium, only some sophisticated models include this effect.

Transmission medium in unventilated spaces: Vuorinen et al. [12] were among the first to propose a model with a transmission medium shortly after the outbreak of COVID-19. At the time of writing this dissertation, their article ranks among the most cited compared to other contributions discussed in this section. The authors discuss several modeling techniques, one of which extends a crowd model by a disease transmission

model. In this manner, they simulate virtual persons moving across a generic space and in a fictitious supermarket. Infectious persons release pathogens into the environment, while susceptible persons continuously take in a fraction of the pathogen concentration at their current position. Emitted pathogens instantaneously distribute within a volume of $1\,\mathrm{m}^3$. After that, the dispersion of the pathogen concentration is approximated by the diffusion equation in the horizontal plane. Here, the diffusion equation serves as a surrogate for dispersion. It actually captures the average movement of, in the Brownian sense, randomly moving airborne particles. In the two-dimensional (2D) case, spatiotemporal changes in the pathogen concentration, u, are described by

$$\frac{\partial u}{\partial t} = D\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + R_s, \tag{3.15}$$

where D denotes the diffusivity or diffusion coefficient, and R_s represents source or sink terms. A term for convection or advection, $\nabla(\boldsymbol{v}u)$, would be added on the right side of Eq. 3.15 [210, pp. 51–3]. Thus, one could account for the velocity field $\boldsymbol{v}$ the aerosol particles are moving with. However, the contributions I included in the literature review typically neglect this effect. Vuorinen et al. [12] solve the partial differential equation (Eq. 3.15) with the aid of a finite difference method. They define the sink term as $R_s = -\frac{u}{\tau} = -\frac{u}{100\,\mathrm{s}}$, which translates into an exponential decay of the viral load with a half time of $\ln(2)\,\tau \approx 69\,\mathrm{s}$. A source term accounts for infectious individuals who constantly or sporadically release aerosols around their position. The diffusivity is assumed to be orders of magnitude higher than valid for ordinary diffusion to obtain a discernible spatiotemporal spread of exhaled pathogen concentrations. The diffusion constant is set to $D = 0.05\,\mathrm{m}^2\,\mathrm{s}^{-1}$ [12], although values in the order of $10^{-11}\,\mathrm{m}^2\,\mathrm{s}^{-1}$ or even $10^{-12}\,\mathrm{m}^2\,\mathrm{s}^{-1}$ would be appropriate for the diffusion of aerosolized particles with diameters between $0.1\,\mu\mathrm{m}$ and $10\,\mu\mathrm{m}$ [86, p. 153]. This disagreement is motivated by the fact that Vuorinen et al. [12] and developers of similar models [15, 194, 199] divert the diffusion equation from its intended purpose and capture the dominant transport phenomenon of aerosol dispersion instead. Dispersion acts on entirely different scales than diffusion. The authors do not justify the exact value. Only Kanté et al. [199] and Duives et al. [194] cite a value calculated in another study [211] but without explaining how they bridge the gap between their and the referenced model. Using the diffusion equation may be reasonable if the spatial resolution of aerosol concentrations is only roughly examined. This limitation is often not stated explicitly. Therefore, users unfamiliar with the subject, in the worst case policymakers for whom such models are often developed, might trust a model with a fine spatial discretization of the diffusion too much and derive bold or wrong conclusions. In addition, a high spatial resolution of the diffusion conflicts with the overall granularity of the model. It easily exceeds the precision of the simulated process of generating a particle cloud, including its geometry and position, not to mention aspects such as neglected aerosol propagation in the vertical dimension. Another disadvantage results from increased computational effort. Employing the diffusion equation requires a temporal discretization. The numerical solving scheme limits the time step size and, thus, can adversely affect the computation time. I intentionally choose a simpler transmission model to match the level of granularity of the crowd model [16, 17, 212, 213].

Transmission medium affected by airflows: Mukherjee and Wadhwa [19] aim at incorporating ventilation. They manually define a flow field in a 2D model of a building with several rooms. Thus, a semi-analytic description of airflow patterns enables them to represent a directed spread of respiratory aerosols, ignoring Newton's law of motion applied to fluids.

The motion of Newtonian fluids, which is a simplified model for liquids such as air, is mathematically described by the Navier-Stokes equations. Although exact solutions to the Navier-Stokes equations exist for some specific cases, real-world applications usually require numerical approximations with the aid of CFD simulations. Treating such modeling techniques in depth exceeds the scope of this dissertation. Therefore, I refer the interested reader to [214] for an introduction to computational methods for fluid dynamics. This matter is not trivial, and coupling the Navier-Stokes equations with a crowd model holds additional challenges. For example, the 2D crowd model must be extended by a dimension to fit the three-dimensional (3D) flow problem. Further, each virtual person embodies a moving boundary acting on the flow, and both models' discretization must be aligned.

Such couplings are known from related research areas, for example, fire and evacuation dynamics. The tool Fire Dynamics Simulator, which implements a CFD model of fire-driven fluid flow, was temporarily extended by a module for crowd evacuation dynamics [215]. The fact that this extension has been given up reflects how difficult it is to build and maintain a coupling of two complex and computationally expensive models.[4] A similar type of model is designed for evacuation simulations in the context of chemical attacks or bioterrorism [216]. Although such contributions appear promising at first glance, they cannot readily be transferred to the problem at hand. One decisive difference concerns the properties of the sources of contaminants.

Löhner and Antil [195] develop a disease transmission model to fill this gap. They combine a crowd model with CFD simulations of airborne particle transport. The motion of air is described by the incompressible Navier-Stokes equations (partial differential equations), with buoyancy expressed through the Boussinesq approximation. The spatial discretization of the flow uses linear finite elements. A system of ordinary differential equations defines the interaction between the airflow field and respiratory particles. The movement of virtual persons introduces an additional system of ordinary differential equations and immersed boundary conditions in the flow field. These different problems are coupled bi-directionally with sequential time stepping and solved with various integration schemes. Based on this, several everyday scenarios are analyzed in [195–197]. Unfortunately, the high resolution of this modeling technique comes at prohibitive computational expenses, a generally acknowledged drawback [217]. For example, a simulation of transmission between passengers moving through a passage of a train station requires approximately eight hours of computing time on 2^{10} cores for a simulated time of two minutes [197]. I argue that the simulation would have to cover at least 10 min to 20 min

[4] The evacuation dynamics module has been removed from the Fire Dynamics Simulator with the pull request `https://github.com/firemodels/fds/pull/10089` (accessed on April 2, 2024) for increasingly challenging maintenance issues.

to find critical exposure levels to airborne SARS-CoV-2 in a realistic scenario of disease transmission via airborne respiratory particles. Hence, the computational effort would increase five or ten-fold. It would further rise by orders of magnitude if the aim were to adequately quantify uncertainties in the simulation input and output. As a consequence, this task quickly becomes unfeasible.

Modeling and simulation of aerosol dispersion without moving persons: Some transmission models use CFD simulations but with static instead of moving persons. Respective simulation studies often concentrate on exhalation of respiratory droplets, particle transport, and evaporation processes. For detailed insights into how the fluid dynamics community regards SARS-CoV-2 transmission, I refer to the reviews by Sheikhnejad et al. [218] and Sedighi et al. [217]. Here, I focus on the travel distances of exhaled particles because this information backs my simulation studies in Part II.

Xie et al. [105] predict the horizontal travel distances of a respiratory droplet for varying conditions. Aerosol particles with diameters up to 200 µm emitted through normal breathing are carried less than 1 m. Another study concludes that 1 µm aerosol particles travel about 0.3 m given that a steady state air jet flow forms through constant breathing [107]. Vuorinen et al. [12] consider a less isolated setting and run CFD simulations of a coughing person in a large room with ventilation. They concede that ventilation in indoor environments leads to turbulent conditions and, hence, substantial variance in the realizations of the aerosol spread. Such uncertainties become more relevant the more realistic the scenario is, for instance, in a reenacted choir practice scenario [219]. The study demonstrates how thermal plumes caused by the singers and radiators in the room influence turbulent flow patterns. Aerosol dispersion differs notably depending on whether buoyancy is considered.

These findings emphasize that there are limits to how far CFD simulations can be taken. They serve well to scrutinize controlled small-scale scenarios, for example, a single exhalation or airborne transmission between two persons positioned face to face. The closer the scenario is to realistic daily-life situations, the more uncertainties must be treated. This, in turn, involves repetitive simulations and conflicts with limited computational resources.

3.2.2.5 Further common aspects

Apart from the above discussion of the basic concepts of modeling inter-individual transmission, I identified further aspects that are treated similarly in the reviewed publications (Table C.2).

Modeling the individual degree of exposure: The model output, particularly in the transmission medium models, is the individual exposure defined as the inhaled dose measured in number of particles. This quantity is typically the accumulated product of an average pulmonary ventilation rate and the number density of aerosol or pathogen particles. Additionally, some contributions include masks with varying filter efficiency,

E_f, simply by multiplying the shedding or intake rate with the term $(1 - E_f)$ [178, 188, 194]. If the filter efficiency is considered to be equal for the whole population and does not affect aerosol propagation, the simulation result is only scaled linearly.

Based on the individual degree of exposure, several models (see Table C.2, marked with S or $E \to I$) employ a dose-response relationship to estimate the number of infected. In some cases, the dose-response is realized merely by a threshold exposure above which any person is considered infected, although this does not reflect reality. A more sophisticated approach relying on an exponential function, as with any Wells-Riley-like model, leads to problems with the calibration and validation because data is scarce. As a consequence, obtaining reliable estimates for absolute infection risks is difficult, if not impossible. Contrarily, evaluating individual exposure levels for several scenarios and ranking the respective exposure levels appears practicable and less controversial.

Modeling and simulating everyday situations – typical case studies: Almost every publication listed in Table C.2 presents not only a disease transmission model but also a case study for a specific scenario. These scenarios are usually either relatively featureless or tailored to a specific situation in everyday life. In featureless scenarios, a virtual crowd moves across an unspecified rectangular area. This is often paired with an unrealistic parameter set for human movement, sometimes assuming even random movement. The knowledge gained from such simulations can, therefore, only be of a general nature. Furthermore, such scenarios fail to exhaust the potential of employing crowd models to mimic realistic everyday situations.

In contrast, the specific scenarios represent clear-cut circumstances. This starts with determining the pathogen to be considered. If SARS-CoV-2 is not the subject, it is influenza virus, SARS-CoV, or Ebola virus [167–172]. In case of SARS-CoV-2, the setting is occasionally outdoors [11, 165, 180, 182, 183, 191, 192] but largely indoors. As discussed in Section 2.1.2, outdoor situations are relevant for short-range transmission via larger respiratory particles, while long-range transmission through smaller respiratory particles is associated with indoor scenarios. The difficulty is to clarify which particle sizes are small and which are large. This is often defined only qualitatively. Most indoor scenarios concentrate on situations with potentially high infection risks. That is, people are present for at least several minutes, and crowd densities are relatively high. This explains partly why simulations often target activities in restaurants or bars and at workplaces such as offices, universities, or construction sites. Also, settings in retail stores or public transport are frequently analyzed. Simulations of pedestrians arriving at an airport gate or boarding and deboarding an airplane receive exceptional attention [167–171, 175, 185, 186, 188, 196, 203, 205–208].

The specific scenarios sometimes boil down to basic movement patterns, such as people forming a line or moving in the opposite direction. Ying and O'Clery [181] locate the highest exposure times in a supermarket near the checkout, where customers usually queue. This is a plausible reason, giving many authors reason for simulating queues in various settings [13, 17, 170–172, 180, 197, 213]. Derjany et al. [172] find that queue formation considerably influences the infection risk. Garcia et al. [180] identify a linear

queue with physical distancing as least critical compared to denser configurations with a winding queue and small distances between the rows. However, this estimate is based on transmission via proximity and, therefore, somewhat obvious because the linear queue exhibits the lowest density and contact times.

Utilization of crowd models: The studies in my review typically employ established locomotion models to mimic human movement. In principle, any crowd model can be extended by a module for disease transmission. Most authors use an adaptation of the so-called social force model (SFM). Alternatively, the movement is determined by real data, so virtual persons follow trajectories extracted from pedestrian tracking data [165, 180, 191]. Few authors define the movement through a cellular automaton [13, 205–208], and, surprisingly, some set up new, non-validated rules or random movement [12, 187, 190, 198–202], although sophisticated and scientifically approved alternatives exist. Developing a new model and simulation framework is prone to errors. Therefore, I argue that corroborated and advanced models should be preferred, especially since validated and verified open-source software is already available. For example, the frameworks Vadere [220] or JuPedSim [221] offer state-of-the-art crowd models. The following excursus introduces the theoretical background of such models very briefly.

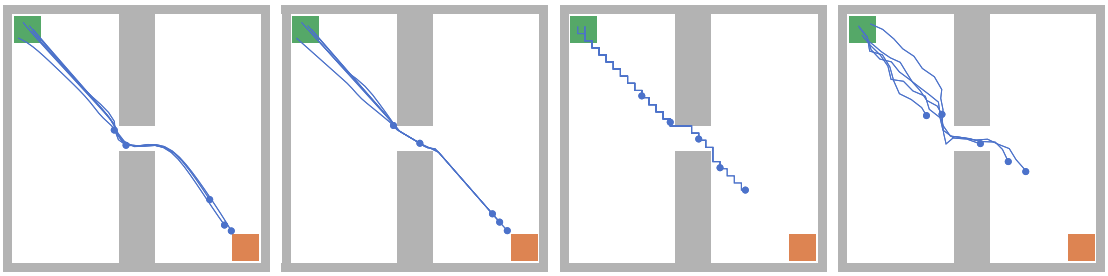
3.2.3 Excursus: microscopic crowd models

Several crowd models have been developed over the last decades to study human movement in various environments and, ultimately, to improve the safety and comfort of crowds in traffic situations. Important approaches and their classification have often been reviewed, for example, in [222–224]. Therefore, I cover solely the basics of so-called microscopic crowd models.

Here, the term microscopic means that virtual persons are regarded as individuals. The spatial and temporal resolution is in the order of centimeters and a few hundred milliseconds, respectively [224]. Note that this resolution usually does not go beyond an individual's position or velocity. Nonetheless, a few disease transmission models evaluate exposure as a function of the head orientation [14, 204, 209]. Presumably, such information can be obtained from the pedestrians' heading direction. However, the heading direction is not a standard model output [225]. In addition, these submicroscopic details have yet to be validated and verified.

Hoogendoorn and Bovy [226] propose to split pedestrian behavior into three layers one of which is denoted as the operational level. It captures a pedestrian's interaction with the environment while walking, as visualized for four microscopic locomotion models in Fig. 3.5. A microscopic approach characterizes pedestrian movement usually by information about the individual positions and states, for example, the velocities, over time.

The class of force- or acceleration-based models (see Fig. 3.5a) treats these quantities in a continuous manner. In 1975, Hirai and Tarui [227] made an attempt to characterize the evacuation of a crowd through forces. Helbing and Molnár [228] introduced the social force model (SFM) twenty years later. Their seminal work has been adapted in various ways, for instance, in [229–236]. Generally, each pedestrian is treated as pointlike mass,



(a) Acceleration-based **(b)** Velocity-based **(c)** Position-based, discrete space **(d)** Position-based, continuous space

Figure 3.5: Comparison of microscopic crowd models: Virtual persons (blue dots) enter the simulation through a source (green) and navigate around obstacles (gray) to reach the target (orange). The paths they choose (blue trajectories) in (a) to (d) depend on the variable chosen to characterize human locomotion and on how interactions between pedestrians are modeled. Note: own figures created with Vadere [220]; see Appendix E (DS3) for simulation files.

obeying Newton's laws of motion. Attracting and repulsive forces act on the mass and, thus, accelerate or decelerate a pedestrian. These *social* forces are rooted in Lewin's [237] theory of social fields, which govern behavioral changes. The respective mathematical formulation is a set of second order ordinary differential equations. Chraibi et al. [238] attest force-based models the ability to reproduce collective phenomena observed in human crowds. However, the differential equations can provoke unrealistic oscillating trajectories and overlapping pedestrians mostly because of inertia [232, 238–240]. These artifacts could impede accurate modeling of locomotion patterns such as physical distancing [225].

Velocity-based models (see Fig. 3.5b) eliminate this problem related to inertia. They directly alter the direction of motion as a function of a pedestrian's environment. This function aims at avoiding collisions. To that end, it typically takes into account the distance to neighboring pedestrians and their velocity. The resulting ordinary differential equations are of first order. Discretization can lead to oscillating trajectories under certain conditions, but more sophisticated models circumvent such limitations [222].

In contrast to the continuous approaches, cellular automata models (see Fig. 3.5c) discretize pedestrian movement. The space is divided into grid cells. A pedestrian's step from the currently occupied cell to a free cell is determined at discrete times by predefined rules. Differential equations are not involved, which is why such models are also referred to as zeroth order models [222]. Gipps and Marksjö [241] present the first cellular automaton for human movement. Burstedde et al. [242] improve the cellular automaton for crowd dynamics. They employ a floor field to define areas of attraction, such as a preferred path around obstacles or crowded places. This idea has been integrated also in other microscopic crowd models to find the direction of motion [243–245]. A common approach represents the floor field by the solution to the Eikonal equation. Figuratively, this yields the shortest path perpendicular to an advancing wave front [245]. The concept of cellular automata is easy to grasp and computationally cheap. However, its capabili-

ties of reproducing realistic movement patterns are limited, and artifacts are a common issue although the original model has been improved multiple times [223, p. 34].

The optimal steps model (OSM), proposed by Seitz and Köster [246] and enhanced in [244], adopts the decision-based approach of the cellular automaton but permits motion in a continuous space (see Fig. 3.5d). The temporal discretization adheres to an event-driven update scheme. I favor the OSM as a basis for my transmission model because of the following and other reasons detailed in Part II. One important advantage is that natural trajectories are obtained simply by optimizing a utility instead of solving differential equations. The event-driven design lends itself to modeling disease transmission because it aligns inherently with respiratory events such as coughing or breathing. Furthermore, the OSM explicitly parametrizes the concepts of intimate and personal space. Consequently, tuning the model to capture physical distancing is straightforward, as demonstrated for scenarios of a bottleneck [247] and a queue [17, 213]. This is particularly relevant for modeling crowds in times of the COVID-19 pandemic. These adaptations rely on the assumption that all persons maintain the same average distance. An informed calibration recognizing experimental and empirical data requires more investigations such as the ones in [248, 249]. However, these uncertainties appear bearable in comparison to other unknowns arising with epidemiological extensions of microscopic crowd models.

3.3 Approaches toward modeling and simulating infectious disease spread at a glance

The preceding sections separately discussed each approach toward infectious disease modeling. Their variety makes it difficult to keep track of the different scopes and resulting tradeoffs. Therefore, I give an overview of typical properties in Table 3.1.

This comparison illustrates a typical conflict of objectives: An increasing degree of detail ensures a more realistic representation of reality while complexity and computational effort rise as well. Nonetheless, one can assume that individual-based exposure models usually exhibit moderate computing times because they operate on much smaller scales in time and space than the large-scale network models.

Each of the categories listed in Table 3.1 comprises many different approaches and probably even more different algorithms and implementations. However, only a few studies disclose the underlying discretization, solving scheme, and program code in detail. Program code is often created and maintained separately for each study or by each research group. However, at least some contributions employ established development environments or simulation frameworks for individual-based modeling and extend them by mechanisms for pathogen transmission. Among these are, for example, the development environment NetLogo [250], the established simulation tools MATSim [151] for transport simulation at large scales and JuPedSim [221] or Vadere [220] for microscopic crowd simulation at local levels.

Table 3.1: The comparison highlights typical scopes, employed methods, and advantages (+) and disadvantages (−) of four model types for disease or pathogen transmission.

Sec. 3.1.1: SIR-like compartmental models predict the spread of a disease typically within larger, homogeneously mixing populations based on differential equations.	
+ Simple to implement and solve	– Strong assumption: homogeneously mixed population
+ Extensible to arbitrary compartments	
+ Low computational effort	– No insights on small populations
Sec. 3.1.2: Epidemiological network models predict the spread of a disease within small to large, heterogeneously mixing populations with the aid of a rule- or equation-based model for disease transmission and progression.	
+ Higher resolution of spreading dynamics compared to SIR-like models	– Realistic networks are difficult to obtain and can be computationally demanding
	– Complex calibration and validation
Sec. 3.2.1: Wells-Riley-like models predict the probability of infection or number of secondary infections for transmission via inhalation in homogeneously contaminated indoor spaces with ventilation rates based on an exponential probability distribution.	
+ Applicable for risk assessment of everyday situations	– Restricted to airborne aerosol particles
	– Limited calibration data (SARS-CoV-2)
+ Low computational effort	
Sec. 3.2.2: Individual-based exposure models predict the individual degree of exposure (or infection probability), considering individual movement patterns, with the aid of a rule- or equation-based model for pathogen (or disease) transmission.	
+ Applicable for risk assessment of everyday situations	– Modeling human movement is complex, but validated models and open-source software exist
+ Simulation results easy to convey to a broad audience	– Limited calibration data (SARS-CoV-2)
± Moderate computational effort (unless aerosol spread is modeled in detail)	

One limitation concerns all models that require or produce information about individuals' exposure and infection states. In case of SARS-CoV-2, they all suffer from insufficient empirical data. This problem can partly be handled by quantifying uncertainties. Once the most influential parameters are known, one can tailor the data collection accordingly to gain more knowledge and make simulations more reliable.

3.4 Summary

A substantial goal of mathematical modeling of infectious diseases is to predict the spread of a disease in a population. Large-scale approaches such as SIR-like systems of differ-

ential equations or individual-based network models capture problem settings on global to national or urban levels. These dimensions are suitable for estimating the course of a pandemic, but they are not intended to gain insights into local outbreaks. The latter is crucial to counteract epidemic or pandemic diseases with transmission dynamics driven by local infection clusters, for example, COVID-19 superspreading events.

In contrast to the large-scale models, small-scale models such as Wells-Riley-like models estimate the infection risk for local scenarios. This risk assessment relies on an exponential probability distribution. The Wells-Riley equation and its adaptations assume well-mixed pathogen concentrations in indoor environments. This assumption is justified for many situations of airborne transmission in rooms with sufficient air mixing. However, there are conditions for which airborne pathogen concentrations are better characterized as inhomogeneous. In these cases, it is necessary to differentiate between areas with elevated exposure risks and lower exposure risks.

The COVID-19 pandemic intensified the efforts to develop models built on the inhomogeneous assumption. A relatively fine resolution of transmission requires that also the movement of individuals is captured, at best, by established microscopic crowd models. These crowd models can be extended by a component for estimating the exposure or, provided that an accurate dose-response relationship can be determined, the infection risk. Such extensions essentially express transmission as a function of the proximity to an infectious person or via a transmission medium, that is, contaminated aerosols or surfaces. Proximity-based approaches fail to capture transmission via aerosols that remain airborne for a prolonged period. A susceptible person can become exposed to airborne pathogens, although not standing near an infectious host. There is a need to address this shortcoming of state-of-the-art models.

Recent developments that are based on a transmission medium seek to remedy this deficiency by explicitly modeling aerosols. In the model, aerosols are often considered to decay exponentially, to spread in the radial direction over time, or to be influenced by airflows. When considering such aspects, it is important to find an appropriate level of granularity to match the crowd model, available knowledge about parameter values, and computational resources. Above all, high-fidelity computational fluid dynamics (CFD) models yield a fine resolution of the transmission medium affected by airflow. However, they entail a multitude of variables, and solving the underlying differential equations results in extraordinary computational costs. Individual-based transmission models that use simpler equations or rules to characterize pathogen spread produce similar information while requiring less information and computing time. Thus, they allow a broad spectrum of users to reenact airborne pathogen transmission in specific everyday situations. Furthermore, the models' parsimony facilitates adequate quantification of uncertainties, which are omnipresent in infectious disease modeling, especially regarding SARS-CoV-2 transmission.

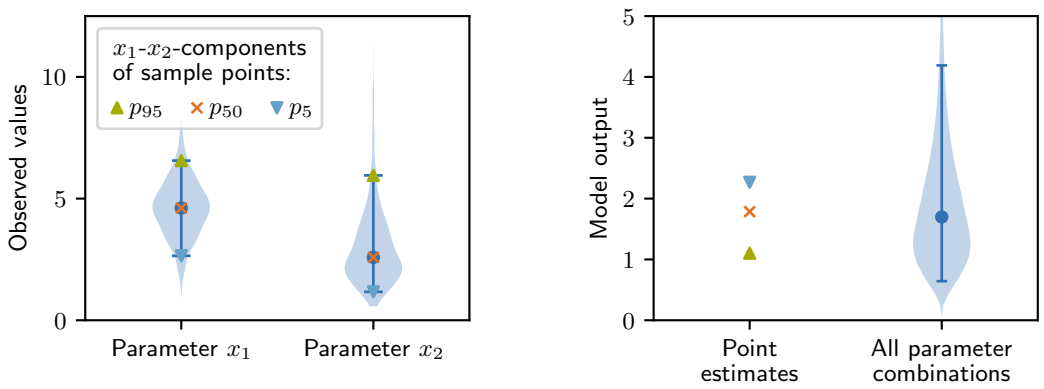
4 Uncertainty quantification

The preceding chapters revealed that there are many sources of uncertainty in experimental observations and modeling of disease transmission. Stochastic simulators represent an important example of how uncertainties get into simulations by incorporating the randomness of a system. Independent of whether models are deterministic or stochastic, they approximate real systems as they are built on simplifying assumptions. This can cause discrepancies between calculated and true values. Errors and uncertainties propagate through the model and diminish the robustness of simulations, especially if small deviations in the input result in large changes in the output. In the words of Neil Ferguson, a renowned mathematical epidemiologist, this emphasizes that "models are not crystal balls" [251, p. 317]. Quantifying uncertainties and analyzing the model sensitivity regarding the input are essential aspects of scientific computing to improve the credibility of predictions [252, p. 12]. Nonetheless, they are often ignored. Funk et al. [253] state that it is uncommon in infectious disease modeling to properly analyze uncertainties. Gödel et al. [254] encounter similar conditions in the field of crowd dynamics.

Neglecting uncertainties means that many simulation studies rely on estimates for a single point in the parameter space. Two major advantages of point estimates are that they require comparatively little effort, and it is easier to communicate compressed information to decision-makers. However, they can give a wrong sense of certainty. It may seem natural that average input values lead to average results and lower or upper bounds of a possible input yield minimum or maximum outputs, respectively. These presumptions do not apply generally, as can be concluded from the following example I adapted from [62, p. 326]. Fig. 4.1 visualizes uncertain inputs and outputs of a model. The mathematical formulation is assumed to be unknown or inaccessible, and the input parameters are independent. The model output in Fig. 4.1b shows that point estimates can be misleading since using single average or extreme values as input (p_5, p_{50}, or p_{95}) does not return the actual average output or extreme values of possible outcomes.

Fortunately, there are techniques for putting vague results of a single simulation run on a firmer footing. As an example, they allow us to assess the model output, also termed quantity of interest, statistically by a confidence interval. In addition to quantifying the probability of output values, investigating the model sensitivity is often helpful. That is, one identifies the input parameters with substantial impact on the output to gain a better understanding of the model or to improve the respective input uncertainty in a subsequent step. This chapter describes how to obtain such quantitative information with the aid of uncertainty quantification methods.

It starts with a definition and classification of uncertainties, followed by an introduction to metrics of model sensitivity and forward propagation. I briefly discuss how case studies in infectious disease modeling approach the task of uncertainty quantification.



(a) Uncertainty in the input

(b) Uncertainty in the output

Figure 4.1: The uncertainties in the input parameters to a model, $f(x_1, x_2) = \frac{x_1}{x_2}$, are characterized by a normal distribution (x_1) and log-normal distribution (x_2). The model output in (b) shows that estimates for the points p_5, p_{50}, and p_{95} yield different results than all possible parameter combinations. Error bars represent the 5^{th} and 95^{th} percentile around the median (circle).

An overview of relevant uncertainty quantification tools and frameworks completes this chapter.

4.1 Definition and types of uncertainty

One way of approaching the concept of uncertainty is to categorize it into two types. A common taxonomy differentiates between aleatoric and epistemic uncertainty. Aleatoric uncertainty, derived from the Latin word *alea*, refers to uncertainty arising from the randomness of the observed system. This type of uncertainty cannot be reduced, unless the nature of the system is altered. In contrast, epistemic uncertainty, coming from Greek *epistēmē* and translated as knowledge, occurs because of a lack of knowledge. This deficiency is clearly an attribute of the observer and can be reduced by collecting information [255, p. 4].

A mathematical description of uncertainties often utilizes probability distributions, as in Fig. 4.1. In these terms, probabilistic simulators incorporate aleatory uncertainty by drawing randomly from a distribution with an uncertain outcome. Epistemic uncertainty would imply, for example, that it is unclear which distribution describes a parameter best. Epistemic uncertainty can be classified into further subcategories of model form and parametric uncertainty [255, p. 4].

The model form is always uncertain to some extent because discrepancies between the model and the observed system are inevitable. For instance, a model excludes variables, such as the variability of airborne pathogen concentrations in indoor settings. Such deliberate decisions are necessary to reduce the complexity of the problem. The consequences of such differences between model and reality often become apparent when the model is

validated against a ground truth. Deviations of the computed value from the true value are typically recognized as errors [256]. There remains doubt about the correctness of the model for the codomain where true values are not available.

Parametric uncertainty entails doubts about the values chosen for specific model parameters [255, p. 4]. It often originates from insufficiently accurate measurements. This is a major source of uncertainty in the context of modeling SARS-CoV-2 transmission on local scales. For this reason, I focus on quantifying parametric uncertainty and its effect on the simulation result.

Regardless of what exactly causes epistemic uncertainties, accumulating more knowledge is usually costly. Therefore, it is advisable to first identify areas where small actions lead to major improvements. Probability theory offers several useful methods to carry out this task.

4.2 Uncertainty quantification methods

There is a variety of uncertainty quantification methods available for different problems. Two classical categories are forward and inverse problems. Forward propagation refers to the practice of pushing uncertainties, for example, originating from poor knowledge about parameter values through the considered system. The resulting variability in the quantity of interest is typically quantified by summary statistics or a probability density function. Such uncertainty analyses are crucial to evaluate the likelihood of specific predictions. Additionally, one can put the input and output uncertainty into relation and derive the model sensitivity. Sensitivity analysis is sometimes classified separately alongside forward and inverse problems. In accordance with [255], I regard sensitivity analysis as one aspect of forward uncertainty problems. In contrast to forward propagation, inverse techniques aim to infer the input parameters upon the condition that the model returns a specific output. Inverse methods are only helpful if the model is generally understood and the database is qualitatively and quantitatively adequate. This does not apply to the use case of this dissertation because data for modeling SARS-CoV-2 transmission is insufficient. Instead, I focus on forward problems, more precisely, sensitivity and uncertainty analysis.

4.2.1 Global sensitivity analysis with the Sobol' method

Sensitivity analysis aims to quantify the impact of variations in the model input on the quantity of interest. An important goal is called factor fixing. That is, one excludes negligible sources of uncertainty and, thus, reduces the size of the problem. In contrast, having identified the most influential parameters allows us to reduce the output uncertainty effectively if a true value can be determined for the parameters with the highest priority. I refer to [257, pp. 49–56] for more details about settings of sensitivity analyses.

A common classification of sensitivity analysis distinguishes between local and global methods. Local methods assess the effect of small perturbations around a reference value. The one-factor-at-a-time method is a popular local screening method. As the name suggests, only one parameter is varied at once to observe changes in the output. However, this does not reveal interactions between variables, which are often considerable.

This limitation does not apply to global methods. They come into play if the aim is to analyze the impact of parameter interactions or to consider the whole range of input uncertainty [258, pp. 303–4].

Global methods are further classified as screening-, regression-, or variance-based methods. Screening-based techniques, with the Morris method [259] as a prominent example, can return a ranking of the parameter importance. This ranking does not reveal the quantitative differences in the impact of different parameters [258, p. 331]. Regression methods, such as partial correlation coefficients, are suitable for linear models. They can also produce meaningful results for nonlinear models if rank transformation is applied, for example, by partial rank correlation coefficients. However, rank transformation fails to linearize the problem in case of nonmonotonic models [260]. In contrast, variance-based methods are model-free. This means they deliver accurate sensitivity measures regardless of whether the model is nonlinear or nonmonotonic [258, p. 328]. They treat the output variability as a sum of the variance caused by the input parameters. Therefore, they are also called analysis of variance (ANOVA) techniques. I employ an established variance-based method, namely sensitivity indices introduced by Sobol' [261] as cited in [263]. Well-tested implementations of this technique are already available.

Adopting the notation in [264], Sobol' indices are defined as follows. Consider the square-integrable function

$$y = f(x), \tag{4.1}$$

which is defined in the n-dimensional unit hypercube I^n. The function describes a nonlinear model with input $x = (x_1, \ldots, x_n)$ and scalar output y. Partitioning $f(x)$ into a set of orthogonal functions yields the respective ANOVA representation with 2^n summands:[1]

$$\begin{aligned} f(x) &= f_0 + \sum_{s=1}^{n} \sum_{i_1<\ldots<i_s}^{n} f_{i_1\ldots i_s}(x_{i_s}, \ldots, x_{i_s}) \\ &= f_0 + \sum_i f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + \ldots + f_{12\ldots n}(x_1, x_2, \ldots, x_n) \end{aligned} \tag{4.2}$$

The constant f_0 quantifies the mean output, $f_0 = \mathbb{E}[y]$. First-order terms $f_i(x_i)$ are equivalent to the contributions by each parameter x_i, whereas the multivariate functions $f_{i_1\ldots i_s}(x_{i_1}, \ldots, x_{i_s})$ are related to interactions of $s = 2, \ldots, n$ parameters [258, p. 289].

The total variance D of the model output is calculated by squaring Eq. 4.2 and integrating from 0 to 1 over I^n. Similarly as in Eq. 4.2, the variance can be decomposed into the sum of partial variances $D_{i_1\ldots i_s}$ [264]:

$$\begin{aligned} D = V[y] = \int_{I^n} f^2(x)\,\mathrm{d}x - f_0^2 &= \sum_{s=1}^{n} \sum_{i_1<\ldots<i_s}^{n} D_{i_1\ldots i_s} \\ &= \sum_i D_i + \sum_{i<j} D_{ij} + \ldots + D_{12\ldots n} \end{aligned} \tag{4.3}$$

[1] The expression $\sum_{i<j}$ is often used in this context. Here and in the following, it denotes $\sum_{i=1}^{n} \sum_{j=i+1}^{n}$.

Dividing Eq. 4.3 on both sides by the total variance D delivers Sobol' indices of first (S_i), second (S_{ij}), and higher orders ($S_{i_1 \ldots i_s}$):

$$\sum_i \underbrace{\frac{D_i}{D}}_{=S_i} + \sum_{i<j} \underbrace{\frac{D_{ij}}{D}}_{=S_{ij}} + \ldots + \underbrace{\frac{D_{12\ldots n}}{D}}_{=S_{12\ldots n}} = 1 \tag{4.4}$$

For high-dimensional parameter spaces, that is, large n, the number of summands in Eq. 4.2-4.4 becomes impractical. This motivates, firstly, to truncate the exact model representation in Eq. 4.2 after the bivariate functions:

$$f(x) \approx f_0 + \sum_i f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) \tag{4.5}$$

Second-order approximations are usually sufficiently accurate [258, p. 289]. Secondly, it appears useful to bundle the information in a so-called total sensitivity index. The total sensitivity index incorporates all partial variances where parameter x_i plays a role. Hence, we have

$$S_{T_i} = S_i + \sum_j S_{ij} + \sum_{j<k} S_{ijk} + \ldots + S_{12\ldots n}, \tag{4.6}$$

potentially approximated by the truncated representation as in Eq. 4.5 [258, p. 324].

In practice, the sensitivity indices are usually estimated, for example, with the aid of sampling-based techniques. Efficient algorithms to conduct this task are already available and described in Section 4.4.

4.2.2 Uncertainty analysis with the Monte Carlo method

The goal of uncertainty analysis, in the following also referred to as forward propagation, is to investigate and quantify the uncertainty in the model response and, thus, to increase the reliability of associated interpretations. The literature offers a variety of methods for uncertainty analysis. Which one is suitable typically depends on the requirements regarding the level of accuracy and on available computational resources. According to Smith [258, pp. 187–97] and Xiu [265, pp. 3–4], a common categorization distinguishes between:

- ✗ **Direct evaluation methods** construct the mean and variance of the response directly from statistical moments of the input parameters. Hence, they do not return the density of the model response. They are limited to models with linear parameter dependence [258, pp. 188–91]. Because of this restriction, they cannot be applied to my pathogen transmission model.
- ✗ **Perturbation methods** approximate the model response of nonlinearly parameterized models by Taylor series. In practice, the expansion is often truncated at the second order because higher orders are expensive to determine. It produces satisfactory results only for small input and output uncertainties [265, pp. 3, 55–6]. Given the large uncertainties in modeling SARS-CoV-2 transmission, perturbation methods seem ineligible for my purposes.

✓ **Sampling-based techniques**, with the Monte Carlo method being a popular example, are often the method of choice for nonlinear problems. They are intuitive and require only three simple steps: One merely has to create random inputs fitting the joint probability distribution of the parameter space, evaluate the model for these realizations, and describe the respective output statistically. Estimates for statistical moments converge to the true values as the number of realizations, N, approaches infinity. The convergence is independent of the dimension of the parameter space, but the expected value converges slowly at rate $\frac{1}{\sqrt{N}}$. More sophisticated sampling schemes such as the Latin hypercube sampling or quasi-random sequences (for example, the Sobol' sequence [266]) accelerate the convergence of the Monte Carlo method. My model is computationally cheap enough for these techniques to be feasible. Therefore, I use this forward propagation technique.

◦ **Spectral representations** can circumvent the problem of computationally demanding models. They essentially employ polynomial expansions as surrogate models from which one can determine the output uncertainty. However, this raises the question of how to construct the surrogate such that the discrepancy to the original model is acceptable [258, pp. 207–37]. I avoid methods that introduce additional uncertainties.

Weighing the advantages and limitations of these approaches, I select the Monte Carlo method. This is a common choice for similar use cases, as shown in the next section.

4.3 How uncertainties are commonly treated in simulations of SARS-CoV-2 transmission

While uncertainty quantification methods are established practice in some fields of research, they find their way only gradually into infectious disease modeling, particularly in individual-based simulations of SARS-CoV-2 transmission. In this section, I discuss a representative selection of publications that attempt to examine the model sensitivity and uncertainty. In this manner, I complement my literature review about models for inhomogeneous exposure risks (see Section 3.2.2).

Adopting a straightforward approach, Kanté et al. [199] quantify the output uncertainty of a stochastic model based on repetitive simulations. Unfortunately, they do not state the number of repetitions. Hence, the statistical summary remains vague. The authors also investigate the changes in the model output with a one-at-a-time analysis, where one experiment comprises several simulations with a single parameter being varied. They run separate experiments for a few parameters, such as the diffusion rate for aerosol propagation, to compare their effect on the infection risk. Salmenjoki et al. [15] use the same method to analyze the isolated impact of different parameters on the number of persons with a critical degree of exposure. Both studies [15, 199] yield qualitatively plausible results. For example, increased pathogen emission and slower decay of the airborne pathogen load result in higher infection risks [15], whereas masks or social distancing can reduce the infection risk [199]. However, the studies exhibit methodological limitations.

Comparing the influence of multiple parameters with one-at-a-time experiments can be rather confusing, and the method is not capable of detecting the impact of interacting parameters on the output.

Qiao et al. [192] use a factorial design in which they vary five parameters simultaneously. They examine workers on a construction site and their compliance with regulations on face covering, physical distancing, vaccination, ventilation, and isolation. Thus, they aim to identify effective control strategies. According to the authors, their model is yet in the state of a proof of concept. Furthermore, the samples generated from the parameter space are sparse. Consequently, there is still uncertainty about the outcomes in between the sampled parameter values. Optimal or just good enough combinations of measures against transmission could easily be overlooked.

A similar methodology is pursued in [183] and [185] to vary three and two parameters, respectively. In [183], each of the parameter combinations is evaluated 200 times in a Monte Carlo setting to account for randomness in the model and to obtain robust mean values for the overall exposure. However, the spread of the output data is not reported. Namilae et al. [185] compute partial rank correlation coefficients in addition to their factorial experiment to estimate the sensitivity of the output with respect to the two parameters.

D'Orazio et al. [178, 179] conduct thorough sensitivity analyses with the aid of Sobol' indices. This method requires many simulation runs, in the considered contributions several ten thousand [178, 179]. In [179], two consecutive analyses reduce the computational effort. The first analysis is to sort out barely or non-influential parameters, and the second one is to refine the analysis only for influential ones. In either paper, the authors assume that the parameters follow a uniform distribution, which is often seen as a reasonable choice if further information is not available. The sensitivity analyses show explainable trends for the respective scenarios.

Such case studies demonstrate how uncertainty quantification can improve our understanding of a model and support the reliability of simulations. Most publications report reasonable results, but they are barely comparable. They depend much on the scope of the underlying model, on which parameters are investigated, and on the respective parameter ranges. Regarding the choice of the method, Sobol' indices seem promising. Also Monte Carlo simulations can be useful for quantifying uncertainties in the model output resulting from lack of knowledge or the model's inherent randomness. This is in accordance with analyses conducted for microscopic crowd simulation without disease transmission [254, 267].

4.4 Tools and frameworks

Software for uncertainty quantification is already available in programming languages such as C/C++ and Python. It is usually advisable to employ established, well-tested software instead of implementing an algorithm from scratch. To name a few generic frameworks, Dakota [268] provides a flexible and extendable toolbox in C++ for optimization, sensitivity analysis, calibration, and other objectives. OpenTURNS [269] is

a platform for uncertainty quantification by probabilistic methods. The Python library chaospy [270] delivers routines for uncertainty and sensitivity analysis based on spectral representations and advanced Monte Carlo methods, while SALib [271, 272] specializes on sensitivity analysis. For further reading about mature uncertainty quantification software, I refer to [258, 270, 273, 274]. In Chapter 8, I conduct numerical experiments with two open-source Python libraries, SALib and chaospy.

Some uncertainty quantification frameworks have been developed for specific applications, including the crowd simulation software Vadere. Tools tailored to Vadere seem a reasonable choice since I use it in the context of this dissertation. Above all, Gödel's [267] uncertainty quantification framework provides routines for Bayesian inversion, forward propagation, and sensitivity analysis. It is partly based on features of SALib and chaospy. Unfortunately, it has not been maintained actively, nor has it been developed further since its release, so the dependencies on SALib are not up-to-date. Another tool, the EMAworkbench [275], embeds some SALib functions and offers a connector to Vadere, but there are open questions[2] about how such a connector can be integrated sustainably as a feature of the EMAworkbench. Also, some output formats are not supported. Therefore, it is economical for me to use SALib and chaospy directly. In the following, I explain essential functions used for my numerical experiments in Section 8.

4.4.1 Sobol' sensitivity analysis with SALib

The Python library SALib provides several routines for global sensitivity analysis, including algorithms for computing Sobol' indices. The library is decoupled from the computational model. Therefore, the workflow consists of (1) defining the uncertain parameter space, (2) generating the model input with a sample function, (3) evaluating the model, and (4) applying a function to analyze the outcome. Functions for steps (2) and (4) are implemented in separate modules. The following details refer to SALib 1.4.5.

The sample function[3]

```
SALib.sample.saltelli.sample(...)
```

generates a Sobol' sequence for Sobol' sensitivity analysis based on Saltelli's [276] improvement of the approach developed in [261, 263]. Suppose, we consider a problem with d uncertain parameters and a desired sample size, N. The aim is to determine total and first order sensitivity indices. Constructing the sequence comprises the following steps:

1. Generate a base sample $2N \times d$-matrix with random numbers matching the input distributions. It is recommended to use the Sobol' sequence.

2. Partition this matrix into matrix $\mathbf{A}$, which contains the first N rows, and matrix $\mathbf{B}$, which contains the rows $N+1$ to $2N$.

3. Compose matrices $\mathbf{C}_i$ as copies of $\mathbf{B}$ but with column $i = 1, \ldots, d$ taken from $\mathbf{A}$.

[2] Questions see `https://github.com/quaquel/EMAworkbench/pull/145#issuecomment-1175965854`, accessed on April 2, 2024

[3] Since version 1.4.6, the function is replaced by SALib.sample.sobol.sample(...).

The implementation in SALib incorporates further improvements of these steps. The first points of the Sobol' sequence exhibit repeated values [277]. Owen [278] recommends to drop the first $N_{skip} \geq N$ points to enhance the uniformity of samples. Furthermore, the number of samples N and the number of skipped samples N_{skip} ideally adopt values of a power of 2.

The rows of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}_i$ represent the parameter combinations to be evaluated. Hence, we have to run the model $(2 + d)\,N$ times in total, resulting in a major computational problem if the model is expensive. Therefore, it is useful to parallelize the model evaluation. Once the simulations have been completed, we obtain the respective output vectors $\mathbf{y}_A$, $\mathbf{y}_B$, and $\mathbf{y}_{C_i}$. Vector $\mathbf{y}_{AB}$ denotes the output of matrices $\mathbf{A}$ and $\mathbf{B}$, that is, $\mathbf{y}_{AB} = \left[\begin{smallmatrix}\mathbf{y}_A \\ \mathbf{y}_B\end{smallmatrix}\right]$.

Then, we can use

```
SALib.analyze.sobol.analyze(...)
```

to determine the sensitivity indices. The function delivers numerical estimates for the first order sensitivity indices in Eq. 4.7 according to the best practice described in [279]:

$$S_i \approx \frac{\frac{1}{N}\sum_{j=1}^{N} \mathbf{y}_{B_j}\left(\mathbf{y}_{C_j} - \mathbf{y}_{A_j}\right)}{\frac{1}{2N}\sum_{j=1}^{2N}\left(\mathbf{y}_{AB_j} - \overline{\mathbf{y}}_{AB}\right)^2}. \tag{4.7}$$

Analogously, it calculates the total sensitivity index (Eq. 4.8) as defined in [279, 280]:

$$S_{T_i} \approx \frac{\frac{1}{2}\frac{1}{N}\sum_{j=1}^{N}\left(\mathbf{y}_{A_j} - \mathbf{y}_{C_j}\right)^2}{\frac{1}{2N}\sum_{j=1}^{2N}\left(\mathbf{y}_{AB_j} - \overline{\mathbf{y}}_{AB}\right)^2}. \tag{4.8}$$

I refer to [258, p. 325] for a detailed mapping of the expansion denotation (Eq. 4.4 and Eq. 4.6) and the corresponding expectation and variance interpretation on which the numerical estimators Eq. 4.7 and Eq. 4.8 are based.

If required, the function also determines second order sensitivity indices and a bootstrap confidence interval for each sensitivity index. The bootstrap confidence interval indicates an estimate of accuracy at a given confidence level for the considered sensitivity index. Its computation essentially relies on calculating the sensitivity index for a number of resamples. The exact construction using the moment method is explained in [276, 281].

4.4.2 Monte Carlo sampling with chaospy

The Python toolbox chaospy offers high-level implementations for uncertainty quantification based on spectral representations and Monte Carlo techniques. Its interface allows for fast prototyping, one reason why I employ it for uncertainty analysis in Chapter 8. My simulation studies only require a function for classical Monte Carlo sampling, yet the toolbox offers more sophisticated algorithms. The content of this section refers to chaospy 4.3.10.

The workflow is similar as with SALib. The first step is to define the uncertain parameter space as a joint distribution, followed by generating the model input with a sample function. The function

```
chaospy.distributions.Distribution.sample(...)
```

creates by default (pseudo-)random samples for a given sample size. If the sample size is large and a model evaluation is expensive, one can adopt more efficient sampling schemes and exploit parallelism. The remaining steps are straightforward. One simply evaluates the model, collects the output, and summarizes it statistically.

4.5 Summary

Numerical experiments are often rife with uncertainties. These uncertainties originate either from the inherent randomness of the analyzed system or lack of knowledge, for example, about the true value of an input parameter. They can severely impact the reliability of simulations, as variability in the input parameters propagates through the model and causes variability in the output. Evaluating merely average values can give a wrong sense of certainty, leading to inaccurate interpretations. Instead, one should consider the whole range of possible inputs and outputs, typically in the form of probability distributions or statistical moments.

There is a variety of methods that support uncertainty quantification in different problem settings. Forward problems, where the effects on the output are to be determined for given causes of uncertainty in the input, can be addressed with sensitivity analysis and forward propagation techniques. This chapter provided a brief overview of methods, of which two were explained in more detail: variance-based sensitivity analysis with Sobol' indices and sampling-based uncertainty analysis with Monte Carlo experiments.

Sobol' indices represent a global measure of how sensitive the quantity of interest reacts to changes in the input. They detect the isolated impact of parameters and interaction effects between multiple parameters on the quantity of interest. Furthermore, the Sobol' method is capable of quantifying the parameter importance for arbitrary nonlinear relationships between input and output.

Monte Carlo experiments are often conducted to quantify the variability in the output data for given input uncertainties. Although classical Monte Carlo sampling exhibits slow convergence rates, it is the method of choice for many applications because it is straightforward, approaches the true value with increasing sample size, and is suited for nonlinear models.

For uncertainty quantification in practice, one can choose among several mature open-source toolboxes. The Python packages SALib and chaospy are generic and offer tools and interfaces easy-to-use for sensitivity analysis, forward propagation, and other tasks.

Although theory and software are available, quantifying uncertainties has yet to be established as a standard procedure in infectious disease modeling. Particularly, in simulations of individual-based SARS-CoV-2 transmission, uncertainties are often neglected despite various unknowns. Among those studies that employ uncertainty quantification methods, many exhibit methodological limitations. Nevertheless, a few exceptions demonstrate how global sensitivity analysis and sampling-based uncertainty analysis can improve the reliability of the simulation results.

Part II

Modeling and simulating airborne transmission of SARS-CoV-2

Part I of this dissertation introduced basics about transmission of SARS-CoV-2, models for pathogen transmission, and methods of uncertainty quantification. Individual-based transmission models for respiratory diseases traditionally focus on short-range transmission via larger respiratory droplets. Models for long-range transmission via smaller airborne respiratory particles, as is the case for SARS-CoV-2, have only recently been developed. In particular, approaches that account for human locomotion and airborne pathogen transmission are yet in a development stage.

Therefore, Part II addresses the overarching question:

Research question

How can airborne transmission of pathogens such as SARS-CoV-2 be modeled and simulated?

The following chapters approach the research question as described by the modeling cycle in Fig. 1.2. In Chapter 5, I operationalize the previous observations on SARS-CoV-2 transmission and, thus, derive a mathematical formulation. Here, I seek to develop a model that a broad audience can comprehend. Furthermore, it should be transferable to pathogens that spread in a similar way as SARS-CoV-2. The discretized version of my mathematical description is cast into an algorithm. Chapter 6 covers how I incorporate this algorithm as an extension of the simulation program Vadere and how I verify the implementation accordingly. In Chapter 7, I simulate everyday situations such as persons waiting in a line and predict the individual exposure risk. Additionally, I reenact two superspreading events and compare simulation results with recorded data to validate the model. Since quantifying uncertainties in my simulations is complex, I dedicate an extra chapter to this task. The case studies demonstrate the capabilities of my model, how to apply it in practice, and how to interpret respective simulation results in a meaningful manner.

5 Operationalization: from the real world to an individual-based exposure model

"Never posit pluralities without necessity."
Attributed to William of Ockham [282]

Operationalization characterizes a method of turning non-quantifiable or not directly measurable concepts into observable data or phenomena. In the case of SARS-CoV-2 transmission, the concept of an infected person could be operationalized by indicators such as a degree of exposure. However, this is only helpful to a limited extent because, under ordinary conditions, we cannot directly count the number of viable virus a person inhales. We need to exploit other phenomena to infer exposure or infection risks. This chapter addresses the challenge of determining a mathematical, quantitative description of the mechanisms involved in the transmission process. In other words, I seek to answer the question:

Research question

How can important aspects of inter-individual transmission of SARS-CoV-2 via airborne respiratory aerosols be described mathematically?

It is neither necessary nor possible to perfectly model the real world. Instead, I adopt Occam's razor, also called principle of parsimony,[1] to set apart relevant from non-relevant details.

Before diving into the details, I briefly recap the motivation for the level at which I look at SARS-CoV-2 transmission. As laid out in Part I, the dynamics of the COVID-19 pandemic was driven by superpreading events. Superspreading events usually occur at local scales, that is, at indoor gatherings of maybe a dozen to a few hundred persons. One or several highly infectious persons cause many secondary cases. Resolving the level of individual events or specific indoor settings enables us to gain insights into the characteristics of between-host transmission. Here, microscopic crowd simulation comes in handy because it supplies us with ready-made approaches for simulating crowd movement in everyday situations. We can take models such as the optimal steps model [244, 246] as a foundation and extend it by a module for exposure risk assessment and, if applicable, one for dose-response relationships.

Choosing a microscopic crowd model as foundation entails important concomitants for modeling pathogen transmission. First and foremost, the resolution of the new feature should reasonably fit the level of granularity of the crowd model. Second, the model

[1] Occam's razor refers to a popular principle stated by William of Ockham that gives preference to simplicity [282].

should be compatible with any established microscopic crowd model to allow for flexible adaptations. Third, the model should be comprehensible for a broad audience if it targets similar user groups and purposes as the underlying crowd model. Crowd simulations are frequently conducted by researchers and safety engineers to convey risk assessments to decision-makers. The interdisciplinary task of estimating an exposure risk demands researchers with related backgrounds to be capable of setting the model into operation and generating meaningful and understandable outputs. Other, more general requirements independent of microscopic crowd simulation concern the objectivity. I attempt to exclude personal biases by drawing upon generally accepted knowledge about pathogen transmission.

The content of this chapter is partitioned into two sections. This structure aligns with the subsequent processes of exposure and dose-response. Exposure refers to between-host transmission of pathogens, while dose-response refers to the individual within-host immune response to a certain dose taken in. I focus on the former, the exposure. Since the dose-response relationship is unknown for SARS-CoV-2, I propose a simple placeholder.

In the following, I largely cover ideas that I formulated in the modeling section in my publication [17] and that I developed further in [213]. I am the main author of theses two peer-reviewed publications. Marion Gödel, Gerta Köster, and Gesine Hofinger supported me as co-authors of the article [17]. Gerta Köster and Hans-Joachim Bungartz helped me prepare and publish the follow-up study [213].

5.1 Exposure model

I propose two exposure models. One captures transmission via respiratory aerosols and, hence, belongs to the class of recently emerging transmission medium models. Additionally, I describe a simple alternative that adopts the state-of-the-art approach of transmission via proximity.

Both models have in common that each individual possesses a health status. Similarly to compartmental models, I characterize an individual as *susceptible* or *infectious*. Susceptible persons are healthy. Their initial exposure is zero, but they can take in and accumulate pathogens from the surrounding such that their degree of exposure increases accordingly. The exposure represents a rather theoretical number, which can either be interpreted as inhaled aerosol particles or pathogen particles. Infectious individuals continuously emit pathogens via aerosols.

5.1.1 Exposure model based on transmission via aerosols

The exposure model based on transmission via aerosols incorporates two core concepts: first, shedding and intake defined by a respiratory cycle; second, the idea of an aerosol cloud as a transmission medium, carrying pathogens from one to another person. The corresponding model parameters can be adapted flexibly to pathogens that are transmitted via airborne respiratory particles. In the following, I focus on SARS-CoV-2.

5.1.1.1 Shedding and intake of pathogens through breathing

Shedding and intake of SARS-CoV-2 is linked to respiratory activity. Normal breathing is particularly relevant because it is an inevitable, life-sustaining process that results in aerosols remaining airborne for at least several seconds and, thus, contributes significantly to SARS-CoV-2 transmission. Therefore, I focus on breathing, but the basic idea can be transferred or extended to other respiratory activities that also produce fine aerosol particles. For speaking or coughing, one would probably adapt the emission frequency and incorporate larger aerosol particles following ballistic trajectories.

To capture normal breathing, I assume that each person has a respiratory cycle, which starts at an individual time but always spans the same duration. Inspiration is immediately followed by an equally long period of expiration, as visualized in Fig. 5.1. Hence, any person's respiratory activity is described by the respiratory frequency, f, and the respective period, $T = f^{-1}$. This simplifies the actual physiology, yet to a reasonable extent. The simplification well matches the level of detail of human locomotion in macroscopic crowd models. Infectious persons emit pathogens bound to aerosols during the expiration time. That is, infectious persons generate an aerosol cloud[2] every time they breathe out (see Fig. 5.2).

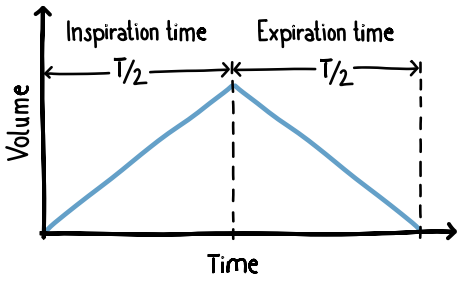


Figure 5.1: The spirometric volume-time curve illustrates schematically how I approximate inspiration and expiration of a single respiratory cycle with period T. The amplitude equals the tidal volume, V_T.

Susceptible persons take in pathogens or aerosol particles during the inspiration time, given that they are currently located within the bounds of at least one aerosol cloud. The number of inhaled particles per inspiration, $N_{P,\,in}$, equals the tidal volume times the particle number concentration in the surrounding:

$$N_{P,\,in} = C_P\, V_T\, (1 - E_f) = C_P\, R_a \tag{5.1}$$

C_P is the cumulative number concentration of particles from superimposed aerosol clouds at the susceptible persons' locations. The tidal volume, V_T, is defined as the volume inhaled and exhaled with every breath [283]. Eq. 5.1 additionally accounts for masks,

[2] The term *aerosol cloud* is inspired by other publications in the field, for example, [111, p. 1]. Aerosol science considers a cloud to exhibit definite boundaries, confining high aerosol concentrations from clean areas [86, p. 379].

which reduce the intake rate by a filtration efficiency E_f. I summarize the tidal volume and the filtration efficiency as absorption rate, R_a, in m^3 per inhalation. Note that respiratory deposition is not modeled in detail since the particle size distribution is unknown. Furthermore, I neglect that respiratory deposition removes particles from the air.

5.1.1.2 Aerosol clouds

Aerosol clouds are defined as follows. Fig. 5.2 visualizes an aerosol cloud as a disk in the horizontal layer at the height of the individuals' heads. This illustration matches the standard 2D perspective of crowd models. The actual extent of the modeled cloud is spherical with radius r and volume $V = \frac{4}{3}r^3\pi$. The cloud shapes during the whole expiration time. Therefore, I place it in the middle between the positions p_1 and p_2 where the infectious person starts and stops exhaling, respectively. The initial shape of the cloud is independent of the walking speed, although it may seem reasonable to use, for example, an ellipsoid for fast walking persons. In this case, the distance walked during the expiration time could be larger than the predefined diameter of the cloud. However, I do not add more complex shapes because, firstly, this would exceed the knowledge about the actual formation of aerosol clouds emitted by moving persons. Secondly, it would not be compatible with other simplifications, such as uniform body heights. Thirdly, the exact contour has little impact since a single cloud contributes only marginally to the cumulative individual exposure, as will become apparent in the simulation studies in Chapters 7 and 8.

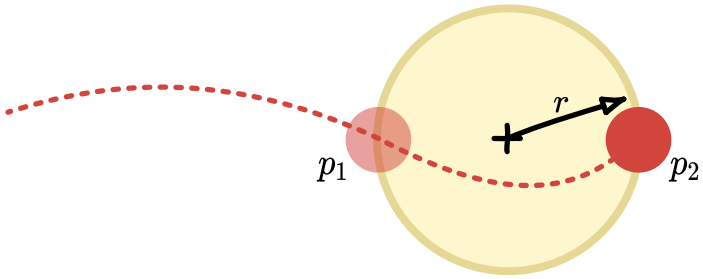


Figure 5.2: An infectious person (red) moves from left to right (dotted line) and exhales an aerosol cloud (yellow disk). The expiration time starts at p_1 and stops at p_2. The aerosol cloud is centered at the midpoint of line segment p_1 to p_2. Its radius r is predefined and does not depend on the movement. Note: own figure adapted from [17, p. 6].

The initial radius r_0, which denotes the radius at the moment when the cloud is generated, is equal for all clouds. Subsequently, the cloud grows linearly over time, t, as defined by a dispersion factor, D:

$$r(t) = r_0 + D\,t \tag{5.2}$$

This simple description of aerosol propagation is based on an experiment and a modeling study described in [87]. Kudryashova et al. [87] recognize (1) a phase of scatter immediately after exhalation, followed by (2) a phase of diffusion. In the second phase, the dominant process is, however, not molecular diffusion but dispersion. Therefore, they employ an atypically large diffusion coefficient to be able to describe this transport phenomenon mathematically. As discussed in Section 3.2.2, diffusion is negligible on the scale of my model, while convective conditions are unknown for the situations that I analyze. Therefore, I focus on scenarios with stagnant air, as can be found more or less in indoor environments without ventilation, and describe phase (1) by instantaneous generation of an aerosol cloud. Phase (2) is approximated by linear growth over time (Eq. 5.2). Since ventilation is not considered, the cloud remains always at its initial position.

Further, I assume that the aerosol particles scatter homogeneously within the cloud. Hence, the number concentration in a single cloud is defined by

$$C_P = \frac{N_P}{V}, \tag{5.3}$$

with the current pathogen load, N_P, and volume, V. The concentration adds up where multiple clouds superimpose. Airborne pathogens are gradually inactivated or removed from the air because aerosol particles evaporate, are carried to higher air levels, or sediment with time. I summarize these effects and assume an exponential decay

$$N_P\left(t\right) = N_{P,\,out}\,\exp\left(-\frac{\ln\left(2\right)\,t}{t_{1/2}}\right), \tag{5.4}$$

where $N_{P,\,out}$ denotes the particles released per breath. If masks need to be considered, $N_{P,\,out}$ is multiplied by the respective mask permeability, similarly as in Eq. 5.1. $N_{P,\,out}$ is constant. I do not include inter-individual differences in the exhalation rate because my simulation studies treat scenarios with only a single superspreader. Furthermore, the simulate time scales are small enough such that intra-individual variability during the course of an infection can be neglected as well. $t_{1/2}$ defines the half-life, that is, the time required for the pathogen load, N_P, to fall to one half of its current value. Once an aerosol cloud reaches less than 1 % of its initial pathogen concentration, I consider it negligible.

I am aware that the above description of aerosol generation and propagation is simple, while real physics of particle clouds are complex. In the eyes of an aerosol researcher, it may appear even simplistic. And yet, I argue that this level of detail is justified for my problem. It attempts to bridge the gap between the established Wells-Riley model, which assumes instantaneously well-mixed air, and models that account for air currents and resolve phenomena of particle transport. Section 3.2.2 underlines the advantages and disadvantages of these two directions. On the one hand, the homogeneous assumption is not well suited for estimating individual exposure risks in rooms with little air mixing. On the other hand, high-fidelity models focus on aspects of particle transport but ignore uncertainties related to aerosol generation, particle size distributions, or respiratory deposition. Numerical experiments conducted with these models are valid for

specific conditions, but they can hardly be generalized. Therefore, such a model would not necessarily provide more useful and accurate solutions than my model.

5.1.1.3 Model parameters for SARS-CoV-2 transmission

The model for shedding, intake, and aerosol propagation has several parameters. They are summarized in Table 5.1. The respective explanation corresponds to the details given in the appendix of my simulation study [213].

Table 5.1: The table provides an overview of model parameters and respective values or ranges.

Parameter	Short description	Symbol	Value / range	Unit
Respiratory cycle period	Time between two subsequent inspirations	T	3	s
Absorption rate	Volume of contaminated air taken in with each breath; equivalent to tidal volume per inhalation if masks are not considered (filtration efficiency $E_f = 0$; see Eq. 5.1)	R_a	$5 \cdot 10^{-4}$	$\frac{\mathrm{m}^3}{\text{inhalation}}$
Initial pathogen load	Hypothetical number of pathogens contained in an aerosol cloud at the instant it is produced (see Eq. 5.4)	$N_{P,\,out}$	10^3	particles
Half-life	Time required for the pathogen load of an aerosol cloud to fall to one half of its initial value (see Eq. 5.4)	$t_{1/2}$	$[6 \cdot 10^2,\, 10^4]$	s
Initial radius	Radius determining the spherical extent of an aerosol cloud at the instant it is produced (see Eq. 5.2)	r_0	$[0.5,\, 1.5]$	m
Dispersion factor	Factor determining unidirectional growth of an aerosol cloud over time (see Eq. 5.2)	D	$[0,\, 6 \cdot 10^{-3}]$	$\mathrm{m\,s^{-1}}$

The respiratory cycle period, T, approximates the average respiratory rate observed in physically active adults of $f = T^{-1} = 19\,\mathrm{min}^{-1}$ [284]. This conforms roughly with the upper limit of the respiratory rate for an adult at rest, which ranges from $7\,\mathrm{min}^{-1}$ to $20\,\mathrm{min}^{-1}$ [285, p. 23].

The absorption rate, R_a, is equal to the tidal volume per inhalation if masks are not considered. The tidal volume in healthy, young adults is typically assumed to be in the order of 0.5 L [285, p. 24]. Here, the exact value is not crucial because the parameter only scales the individual exposure. One could also normalize this parameter to 1, but I argue that values close to measured quantities make it easier to interpret the model setup and results.

This applies in a similar manner to the initial pathogen load, $N_{P,\,out}$, as long as the exposure is not interpreted in absolute terms. The chosen value is somewhat arbitrary but stays in the regime of plausible particles numbers observed for SARS-CoV-2 transmission. As discussed in Section 2.1.1 and shown in Fig. 2.2, RNA emission rates of 10^2 copies to clearly over 10^4 copies per minute are realistic. With approximately 20

breaths per minute, $N_{P,\,out} = 10^3$ copies per exhalation seems a reasonable choice for a highly infectious person.

The half-life of aerosol clouds, $t_{1/2}$, quantifies the time-dependent reduction in the pathogen load due to sedimentation or rise of aerosol particles and deactivation of pathogens. Section 2.1.3 reports a wide range for the persistence of airborne SARS-CoV-2 or aerosol particles, often depending on the ambient conditions and particle size distributions. An analytic analysis estimates half-lives of 5.5 min to 7 min for aerosol particles expelled through coughing and speaking [111]. However, breathing leads to smaller particle sizes (see Fig. 2.4). Smaller aerosol particles remain airborne for longer periods [100]. Therefore, I propose a half-life of about 10 min ($6 \cdot 10^2$ s) as lower limit. The upper limit is given by the stability of airborne SARS-CoV-2. In an experimental setting, the median half-life of viable virus reaches 1.1 h [79]. Another study that takes varying ambient conditions into account derives half-lives of 30 min to 177 min [110]. I use the order of magnitude of the highest value, 10^4 s, as an upper bound.

The parameter initial radius, r_0, quantifies the primary phase of scatter when an aerosol cloud is generated. I derive the range of possible values from numerical experiments. As calculated in a computational fluid dynamics study of a constantly breathing person, 1 µm aerosol particles are displaced about 0.3 m horizontally from the source [107]. Another study accounts for evaporation under several ambient conditions and particles with different sizes and exhalation speeds. Particles with sizes < 200 µm exhaled through breathing are predicted to travel horizontally less than 1 m [105]. These two predictions refer to conditions where the exhaling person stands still. Including normal walking behavior, slightly higher initial radii can be expected. Consequently, I set a range of 0.5 m to 1.5 m. The upper limit conforms with an experiment where aerosol particles of sizes between 0.5 µm and 20 µm are generated artificially. Within a few minutes, the concentration levels off at 1.5 m from the source [87]. Since the modeled phase of initial scatter is much shorter than the observed period in the experiment, my upper boundary seems appropriate.

The dispersion factor, D, determines the spatiotemporal spread of an aerosol cloud. I define a range of 0 to $6 \cdot 10^{-3}\,\mathrm{m\,s^{-1}}$. The lower limit describes the case that the size of the aerosol cloud remains constant. The upper limit is the result of an order-of-magnitude estimate, attempting to answer the question of how far fine airborne particles travel in an unventilated room within a simulated period of 30 min (1800 s). Horizontal displacements of 1 m to 10 m seem possible for this time frame, as can be concluded from simulations of a poorly ventilated restaurant [74] and a Large Eddy Simulation of aerosol spread during a choir rehearsal [219]. Transferring this to my aerosol model, the dispersion factor remains below $\frac{10\,\mathrm{m}}{1800\,\mathrm{s}} \approx 6 \cdot 10^{-3}\,\mathrm{m\,s^{-1}}$. An order of magnitude larger distances of 10^2 m appear unrealistic in view of the analysis by Kudryashova et al. [87]. They disclose that it would take aerosol particles in still air more than 5 min to cover the area 2 m around the source.

5.1.2 Transmission model based on the concept of proximity

The proximity-based approach belongs to the state of the art, unlike the new model described previously. I briefly present one version of how pathogen transmission is determined by two persons' mutual distance. The simulations in Chapters 7 and 8 do not make use of this model because it is not well suited for capturing long-range transmission. Nonetheless, I broach it here to emphasize that there are even coarser approaches attempting to represent pathogen transmission and that my work is not about the ultimate precision in infectious disease modeling but the feasibility of prediction. In addition, the proximity-based model serves as an example in Chapter 6 to clarify how other developers can incorporate alternative approaches into the simulation program.

In the simplest version of the model, any susceptible person becomes exposed when approaching an infectious host so that their distance falls below a defined threshold, for example, 1 m. For a threshold distance of (almost) 0 m, one could speak of transmission via physical contact as defined in [11]. The susceptible person's exposure is either 0 % (false) or 100 % (true). More graduated versions could include an exposure time that needs to pass before close proximity leads to exposure or adopt a stochastic approach with probability distributions determining a chance of exposure. Additionally, the degree of exposure could map to any number in $[0, \infty[$ instead of $\{0, 1\}$ or {false, true}. The consequences of a particular exposure, meaning whether someone becomes ill or stays healthy, are defined separately by a dose-response relationship.

5.2 Dose-response model

The dose- or exposure-response model represents the intra-individual response to a specific dose of pathogens taken in. The relationship is commonly defined by a dose-response curve. The response depends on many factors, including the transmission route. Importantly, as described in Section 2.1, it makes a difference where inhaled SARS-CoV-2 enters the human body and which organs it attacks. In view of the largely unknown biological processes, modeling the immune response and the course of an infection with various possible endpoints from infection to death exceeds the scope of this dissertation. Even focusing solely on early processes and estimating an average infection risk is imponderable. Essential quantities, such as the median infective dose in humans, are unclear (see Section 2.1.1). Therefore, I define a substitute that should be replaced when reliable data and biologically plausible dose-response models are available. At this point, I refer to Chapter 8 in the book by Haas et al. [62] and the discussion in [286], which outline important aspects plausible dose-response models should reflect.

My placeholder model requires an additional health status termed *infected*. The infected state complements the previously introduced statuses *susceptible* and *infectious*. Note that characterizing someone as infected does not inform about whether that person can infect others. That is, the health statuses infected and infectious are not equivalent. Mathematically, I define the infected state in a deterministic way, either as non-infected or infected, which translates into the set $\{0, 1\}$. Alternatively, one could adopt a stochastic approach and specify it as a probability of infection in the range of $[0, 1]$.

I employ a threshold to determine whether a person remains uninfected or becomes infected. Any person with an exposure exceeding this threshold exposure is considered infected. Other recent works, such as in [111, 184, 198], represent the dose-response for SARS-CoV-2 in a similar manner, although thresholds generally do not apply to pathogens [286]. I implement it merely as a placeholder and conduct risk assessments by comparing the exposure occurring in a series of scenarios.

5.3 Summary

In this chapter, I introduced a mathematical description of SARS-CoV-2 transmission via fine respiratory aerosols. My mathematical model targets local scales of between-host transmission in unventilated indoor situations with two to a few hundred moving persons. Human locomotion is captured by established microscopic crowd models, such as the optimal steps model. In accordance with the level of granularity of the crowd model, transmission of COVID-19 is expressed by two separate processes: exposure and dose-response.

My exposure model is based on transmission via aerosols emitted and taken in through normal breathing. Infectious persons exhale a fixed amount of pathogens with every respiratory cycle into the environment. The exhaled breath is represented by aerosol clouds. Susceptible persons inhale a fraction of the airborne pathogens from their surrounding if they are located within one or more overlapping aerosol clouds. These aerosol clouds are modeled as spheres with a predefined radius. They remain stationary at the position where they have been generated, and they can increase over time depending on a dispersion factor. In this manner, the model is able to realize spatiotemporal aerosol propagation accelerated by dispersion. The pathogen load in each cloud distributes instantaneously and homogeneously. It decays exponentially and, hence, has a half-life. The exponential decay mirrors removal of aerosol particles from the height of the persons' heads through sedimentation or time-dependent deactivation of pathogens. Values chosen for the model parameters match SARS-CoV-2 transmission via respiratory aerosols. The parameters can be adapted flexibly if the model is to be applied to pathogens that spread in a similar way. This mathematical description of SARS-CoV-2 transmission is parsimonious. Its simplicity holds limitations, above all, a rough description of aerosol propagation and pathogen concentrations that should not be interpreted in absolute numbers. Nonetheless, the relatively coarse degree of detail is on purpose to match the underlying crowd model. In addition, I introduced a component to describe the infection process of exposed individuals. Until dose-response relationships are attainable for SARS-CoV-2, I use a threshold to indicate infection as placeholder for a biologically plausible relationship. These decisions and all other simplifications avoid giving the impression of false accuracy in the presence of considerable knowledge gaps.

6 Integrating the exposure model into Vadere

"Before software can be reusable it first has to be usable."
Attributed to Ralph Johnson[1]

The next steps in the modeling cycle are to translate the previously derived model into an algorithm and then to cast this algorithm into a simulation program. An algorithm is a sequence of discretized instructions. Once it has been implemented, we can finally execute computations. The implementation is ideally clean, what experienced programmers interpret as "elegant and efficient" [288, p. 7] or "simple and direct" [288, p. 8].

Such definitions apply in particular to software developed to support research purposes. Writing and using research software is a fundamental part of scientific practice across disciplines, as it underpins knowledge gain. Although the quality of research software is crucial, many scientists adopt a relatively informal approach toward software development and maintenance. This originates from insufficient training, little credit paid to developers, or the absence of funding for adequate software engineering [289]. Instead, research software should actually fulfill specific requirements. Barker et al. [290] advocate for applying the principle of findable, accessible, interoperable, and reusable (FAIR) code to make numerical experiments reproducible and sustainable.

Concerning my use case, such considerations raise the question:

Research question

How can the mathematical model be implemented efficiently as sustainable software in the sense of reusable and verified code?

Since little is achieved when everyone creates reusable software but no one re-uses other software, I avoid re-inventing the wheel and build on an established open-source simulation program for microscopic crowd simulation. I choose Vadere [220] because it essentially meets the FAIR criteria and, thus, allows me to create a module for pathogen transmission that also conforms with the FAIR principle as much as possible. My choice is also driven by the fact that I am a member of Vadere's internal development team. This decision is not a manifestation of the *Not Invented Here* syndrome, but I regard it as an advantage to be familiar with the simulation program. It enables me to extend the code faster and better than other freely available crowd simulation tools such as JuPedSim [221].

In the following sections, I document the implementation of my mathematical model as research software. First, I summarize the requirements associated with choosing Vadere and other important prerequisites for developing research software. Then, I explain the

[1] Ralph Johnson organized the first conference on patterns and co-authored the seminal textbook "Design Patterns: Elements of Reusable Object-Oriented Software" [287].

design and the realization of the requirements, including the code verification. This chapter reflects in large parts my ideas presented at the conference on Traffic and Granular Flow 2022 [212]. Gerta Köster co-authored this contribution.

6.1 Requirements specification

In software engineering, requirements determine the software's functional and non-functional properties that stakeholders can expect [291, pp. 455–6]. In the following, I focus on non-functional requirements, while functional requirements are implicitly formulated by the model description in Chapter 5. The stakeholder of my software is, in a broader sense, the scientific community. In a narrower sense, it refers to potential users, scientists from disciplines involved in the problem, for example, physicians, public health researchers, or computer scientists. They should be able to extend the software and to review, reproduce, interpret, or generate new simulation results. This implies several non-functional requirements, which apply generally to research software:

- Findable: Potential users or machines should be able to find the software and identify different versions of it [290]. Vadere including my extension is findable through search engines or direct access to the accompanying research articles [17, 220], the project website,[2] or the source code repository.[3] The latter provides the complete development history and version identifiers.

- Accessible: Stakeholders should be able to access the software [290]. This allows them to develop it further or to run simulations and, eventually, reproduce my numerical experiments. Open-access is the preferred policy since my research is not subject to security issues or other constraints that would necessitate authorization. Therefore, the Vadere code repository including my extension is publicly available.

- Interoperable: The software should be interoperable and, hence, capable of exchanging data or interacting with other software [290]. Chapter 8 delivers an excellent example of how Vadere's interoperability enables me to exploit external packages for sensitivity analysis and forward propagation.

- Reusable: The software must be executable and reusable [290]. This implies that third parties can comprehend it, alter, extend, or integrate it into other software projects. Richards [292] confirms Vadere's usability in general. The reusability of my transmission module is supported by the fact that researchers working for health authorities found the software shortly after its first release and are using it to assess the transmission risk during the boarding process of airplanes.[4]

[2] www.vadere.org, accessed on April 2, 2024

[3] `https://gitlab.lrz.de/vadere/vadere`, accessed on April 2, 2024

[4] Researchers working for the U.S. Federal Aviation Administration's Civil Aerospace Medical Institute, National Research Council Canada, Boeing Company, and Movement Strategies Ltd. contacted me in 2022. Since then, I have technically supported the research consortium simulating pathogen transmission with Vadere (Greenhaw, R., personal correspondence with S. Rahn via e-mail, March 20, 2024).

In addition, there are more specific requirements that I must comply with when implementing my model extension. Vadere is written in the programming language Java [220]. As a matter of course, I adopt the language and the object-oriented programming paradigm. The following concrete requirements are concomitant with Vadere:

- Modular: Vadere is based on the model-view-controller pattern and, thus, adheres to a principle known as separation of concerns [220]. This results in a clean architecture and facilitates altering, extending, or managing a component without affecting another.
- Flexible: Vadere uses human-readable input files [220]. This enables end users without a software engineering background to customize simulations to their needs. Following this approach, also the module for transmission of pathogens should be flexible and adaptable in a simple manner without interfering with the source code.
- Verified: The code should be verified, which means it has been reviewed carefully and appropriately represents the mathematical model [252, p. 146]. Otherwise, the software and numerical experiments cannot be trusted. An accessible quality report contributes to making this step transparent. The Vadere development team has agreed on keeping the percentage of tested lines of code at around 24 % [220]. For our research software, this constitutes an acceptable compromise between flexible enhancements, careful testing, and a limited development budget. Software extensions should generally maintain this coverage.

The following sections document how I realize these requirements. I explain how to approach these criteria, focusing on the aspects of reusable software, most importantly, extensible, modular, flexible, and verified code. In this way, I aim to advance new approaches in research areas such as infectious disease modeling and mass gathering health.

6.2 Embedding the exposure model in Vadere

Extending Vadere by a disease transmission module essentially means that I combine three models:

1. a microscopic crowd model describing human movement by individual trajectories,
2. the exposure model presented in Section 5.1, and
3. the dose-response model provisionally defined in Section 5.2.

Fig. 6.1 visualizes how these three components are linked. They are not mutually dependent since the exposure model requires only the output of the locomotion model, while the dose-response model relies on the calculated exposure level.

In Vadere, the locomotion model or any additionally defined sub-model is updated by a while-loop (see Fig. 6.2). I add the exposure and dose-response model as sub-models to

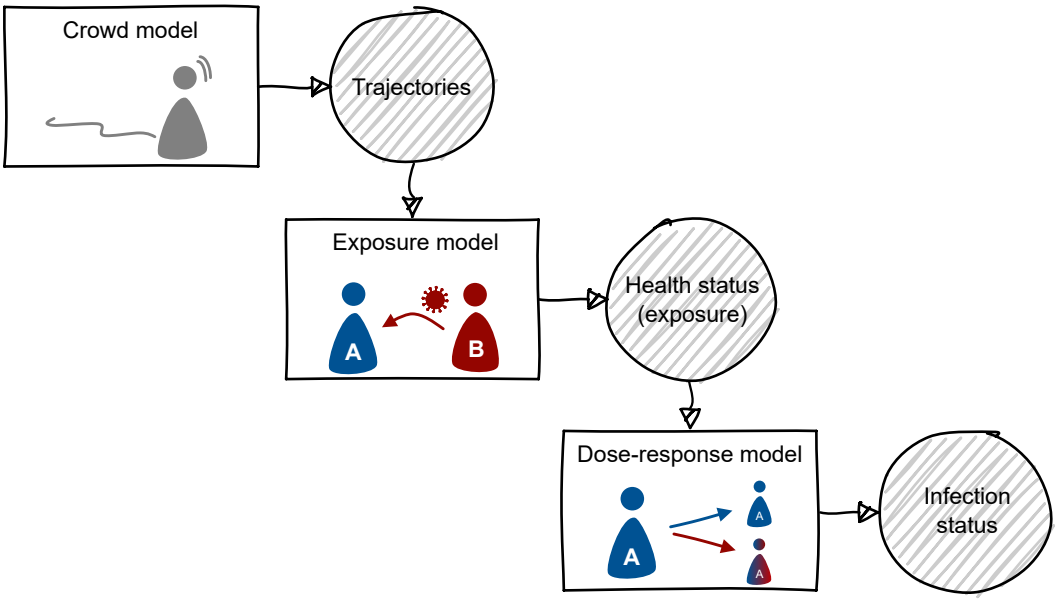


Figure 6.1: The schematic illustration shows the links between microscopic crowd models in Vadere, the new sub-models, and their outputs. Note: own figure adapted from [212, p. 241].

the simulation loop such that they are called and updated independently of other models in every step of the simulation. Consequently, one can combine my extensions with any other locomotion model.

Vadere offers an interface for new (sub)models. The diagram in Fig. 6.3 shows that two separate abstract classes implement this interface and reflect the distinct processes of between-host transmission and dose-response. Abstract classes generalize common behaviors of their subclasses and, thus, improve reusable, extensible, and flexible structures. The abstract classes exist to be inherited by the concrete classes `AirTransmissionModel`, `ProximityExposureModel`, and `ThresholdResponseModel`. These can be instantiated and contain the logic as mathematically formulated in Section 5.1 and Section 5.2, respectively. I include the `ProximityExposureModel` mainly to demonstrate that the structure can easily be extended. This additional example may help developers without a background in computer science to implement alternative sub-models. Other important classes not included in Fig. 6.3 relate to each person's health status and infection status. They are incorporated as wrapper classes in the class `Pedestrian`. Attributes are stored separately from the definition of the model to comply with Vadere's design pattern.

A design pattern represents an abstract, generally applicable solution to recurring design problems. It increases the flexibility and re-usability of an object-oriented software architecture [287]. Vadere exploits the so-called model-view-controller design pattern [220]. As shown in Fig. 6.4, I adopt this pattern and break my code up into three main components: Importantly, the meaning of the term model now shifts compared to previous notations. In the context of the design pattern, it refers to data that describes

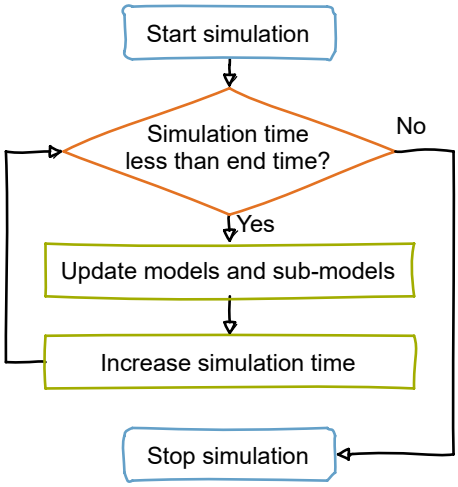


Figure 6.2: A simulation in Vadere is updated by a while-loop. Note: own figure adapted from [17, p. 9].

the present simulation state, whereas the controller is equivalent to what modelers usually understand as a model. That is, the controller contains the rules and instructions I define to capture disease transmission. The view component is responsible for the graphical representation of the data model. It also provides the entry point for users to interact with the simulation program and to alter the attributes of the simulation. I define these attributes as accessible parameters in the model component. This allows users to flexibly tailor the controller behavior to their use case without altering the code. For example, the default parameter set for the `AirTransmissionModel` corresponds to the definition in Section 5.1.1. Reducing the values for the initial pathogen load or half-life of aerosol clouds might suffice to adapt the controller to influenza viruses or other less contagious pathogens than SARS-CoV-2. As an overall result of adhering to the model-view-controller design pattern, I obtain a clean architecture with classes grouped in packages, leaving other parts of the code unchanged.

This separation of concerns entails another benefit, as it breaks up the code into frontend and backend. It enables me to adjust my testing strategy according to separate requirements for the frontend and the backend. The primary purpose of the frontend is to visualize the simulation state. For this reason, it is sufficient to correct visible errors. In contrast, the backend delivers numeric simulation results, which build the basis for further analyses and interpretations. Consequently, the related parts of the software must be tested more carefully to avoid erroneous conclusions.

6.3 Software verification

Verification is a crucial step in the modeling cycle as it improves the reliability of a simulation program. The aim is to ascertain whether the algorithm and software appropriately

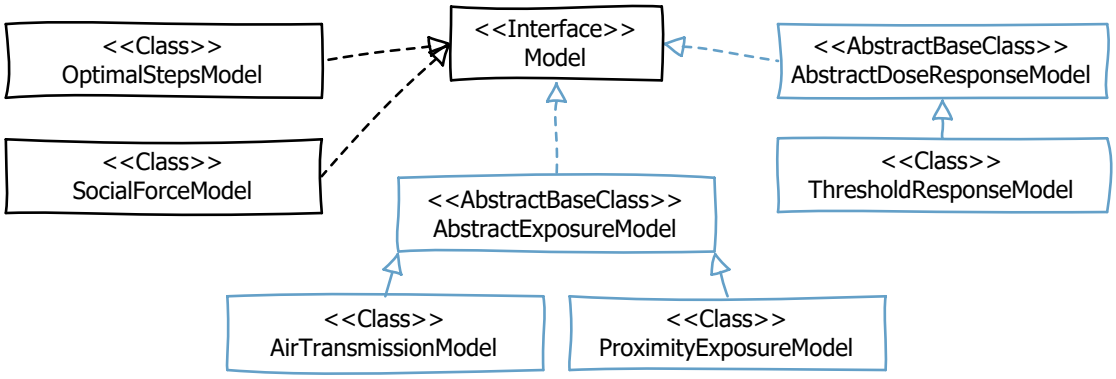


Figure 6.3: The diagram contains essential classes corresponding to the mathematical models for exposure and dose-response (blue) and how I integrate them into the existing structure (black) of Vadere. Note: own figure adapted from [212, p. 242].

represent the mathematical model [23, pp. 14–5]. Deviations of the numerical approximation from the exact solution seem less critical for my use case. The exposure model is coarse, so errors introduced by the discretization appear comparatively small. Therefore, I focus on software verification and report on the software engineering practices that I utilize to minimize the amount of errors in the code.

Prominent techniques for quality assurance include static and dynamic analysis [252, pp. 154–9]: Static analysis, for example, by reviewing and compiling the code, does not involve running the program. In contrast, dynamic analysis requires the code to be executed, as is done within the scope of unit, component, or system testing. Code review or code inspection and testing cannot replace each other but are typically applied in combination [293, pp. 230–2]. Furthermore, while tests are well-suited for detecting errors in the code, they cannot prove the absence of errors [294].

My development strategy toward a simulation program without critical errors encompasses static and dynamic analyses. The static analysis mainly consists of code inspection by other Vadere developers and by a research project partner, who transferred the functionality of my simulation program to a commercial crowd simulation framework named crowd:it [295]. A more formal approach to improving code quality involves testing at different levels. I implemented unit tests at the level of methods and object classes such that the percentage of tested lines of my contributions complies with the line coverage of 24 % in all Vadere modules, excluding the view component. The lines of core classes, such as the `AirTransmissionModel`, are covered at 95 %.[5] Thus, I achieve the verification requirement. The unit tests are complemented by integration tests, which execute the whole simulation program for specified scenarios and check whether changes to the code cause errors in a simulation. The continuous integration pipeline of the software project automatically triggers these tests. It also enforces a standard process for contributing

[5] Coverage report for Vadere 2.1: `https://gitlab.lrz.de/vadere/vadere/-/jobs/3867471`, accessed on April 2, 2024

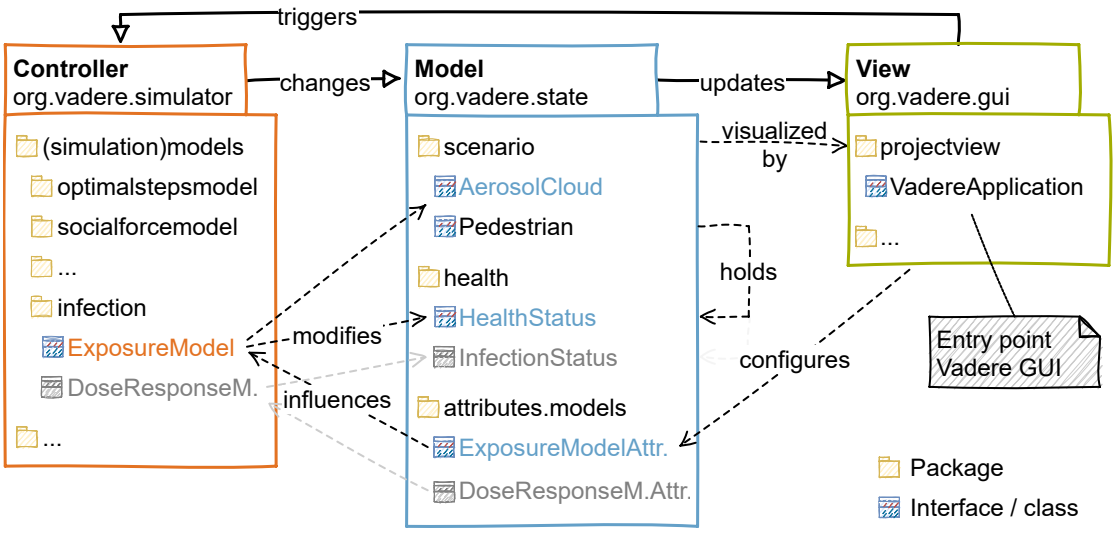


Figure 6.4: The package diagram shows how the components of the exposure model interact. The classes associated with the dose-response model (grayed) adopt the same pattern. Note that, for simplicity, interface or class names used here do not exactly match the code. Note: own figure adapted from [212, p. 244] and [220, p. 21].

to and deploying new releases of Vadere. The whole development process, including the commit history, is laid open to allow others to retrace certain developments and to reproduce the results of an earlier version. As a side effect, the version control system also makes specific releases findable and accessible.

6.4 Summary

This chapter presented how I integrated the previously developed mathematical models for exposure and dose-response into an established tool for microscopic crowd simulation. As a first step, I adopted the criteria of findable, accessible, interoperable, and reusable software that my simulation program should meet. In compliance with these requirements, I selected an established research software for crowd simulation, Vadere, as a foundation for my extension. This choice imposed further specifications regarding the design and testing of my extensions. In addition, I focused on creating a flexible and modular tool.

To this end, I integrated the exposure model and dose-response model as sub-models into the existing structure of Vadere. Above all, the new feature adopts the model-view-controller design pattern, resulting in a modular architecture. The most important part of my model for pathogen transmission via aerosol clouds, corresponding to the description in Section 5.1.1, is embodied in the class `AirTransmissionModel` as part of the controller in the model-view-controller pattern. Corresponding attributes are accessible as input parameters, enabling end users to adapt the controller behavior flexibly without altering

the code. In addition, I incorporated the provisionally defined models for transmission via proximity and dose-response into the software to demonstrate how others can quickly and easily attach alternative transmission models to Vadere.

Quality assurance is an essential step in the development phase of reliable software. My code verification mainly comprised code inspection, unit tests, and integration tests according to standard techniques in software engineering. A continuous integration pipeline supported building, testing, and deploying changes to the software project. In addition, making the code repository public increased transparency and reproducibility. As a result, I equipped Vadere with a reusable and tested module for pathogen transmission in dynamic crowds that can expedite the maturation of infectious disease modeling.

7 Virtual world: simulating everyday situations

"All models are wrong but some are useful."
George Box [296, p. 202]

Now that we have a simulation tool, we can conduct numerical experiments in a virtual world. The aim is to simulate what-if scenarios of pathogen transmission in human crowds. Here, a decisive advantage of simulations comes into play: Virtual experiments do not run into ethical problems. However, the underlying model and resulting outputs can deviate from reality. In a sense, they are always wrong, but the modeler must be concerned only about the essential aspects [297]. If the modeler finds a crucial deviation, the model might not be suitable. For example, a significant discrepancy between the numbers of secondary infections recorded for a superspreading event and high-risk exposure levels observed in the respective simulation would be problematic. Comparing empirical and simulated data is an essential step toward model validation. Thus, one can check whether the model reflects reality reasonably well, given that available empirical data is trustworthy and its level of uncertainty is acceptable [252, pp. 371–2]. However, according to Popper [298, pp. 17–20], it is impossible to conclusively confirm the validity of a scientific statement. Instead, one needs to test the statement through falsification. In my case, this means that if the output does not match empirical data adequately, I must reject the model. If the model passes the test, this step of model validation builds trust. It paves the way for predicting transmission risks in situations for which we currently have no information or experience. Therefore, I now examine the questions:

Research questions

How can one reliably predict the exposure risk for everyday situations?

- How can one validate the model to ensure realistic and meaningful simulation outputs?
- Which relevant scenarios can be simulated with the individual-based exposure model?

To address them, I compare the simulation output and field data of two superspreading events. Then, I systematically select everyday situations of particular relevance and predict the respective exposure risks. Uncertainties are crucial when making such predictions, but this is a more complicated matter. Therefore, I quantify uncertainties and also discuss the simulation results separately in Chapter 8. My modeling paper [17] and parameter study [213] partly cover the simulation results presented here. Since the scientific knowledge about SARS-CoV-2 transmission has grown and my model has undergone an iterative development process between the time of writing the two articles, I now present

an updated version of the validation in [17]. The settings and parameter choices for the renewed validation align with the ones used in [213].

7.1 Validation through reenacting superspreading events

Validating my model through empirical measurements poses a challenge. Most importantly, as explained in the Chapters 2 and 3, all individual-based transmission models for SARS-CoV-2 suffer from insufficient data. Available data sets are typically associated with superspreading events. The disclosed information is often incomplete, for example, due to privacy concerns or because it was unclear which information would be necessary for retrospective analyses. Furthermore, behavioral changes in the population or measures taken during the pandemic affected the transmission dynamics and resulted in biased data. Reconstructions of superspreading events typically rely on surveys and, therefore, involve many uncontrollable variables, not to mention differing standards and lack of truthfulness when reporting official statistics. Finally, these statistics only state the number of secondary cases but not the respective exposure levels, prohibiting direct comparisons of simulated and empirically observed data.

The following approach helps me to overcome these obstacles. I reenact two superspreading events first inspected by Lu et al. [73] and Hamner et al. [129]. These infection clusters occurred during the early phase of the pandemic when mitigation strategies had little or no impact on social gatherings. The reports cannot guarantee that the index patients are responsible for all infections, but it seems likely for the majority of secondary cases. Although particularly the description in [129] is limited for data protection issues, the details given suffice to simulate the two situations. In the absence of a widely accepted dose-response model, I consider any virtual person as potentially infected or exposed at high risk if the exposure reaches the same level as in a benchmark scenario of a close contact. This allows me to compare the model output and the empirical data.

The simulations of the benchmark scenario and the superspreading event rely on the same parameter settings. In the modeling study [17], we analyze only one combination of possible input values. As more knowledge has become available in the meantime, I now use three parameter sets (see Table 7.1) that agree with the settings used in the follow-up study [213]. The updated simulation output is available in Appendix E (DS4). One parameter set consists of the arithmetic mean values of the parameter ranges defined in Table 5.1. The other parameter combinations lead to the maximum aerosol concentration in an aerosol cloud and the maximum spatial spread of an aerosol cloud. Hence, I obtain one moderate sample and two extreme cases. Compared to the parameter set *avg. of sample intervals*, I expect *max. aerosol concentration* to cause higher individual exposure levels but fewer affected persons. *Max. aerosol spread* should produce lower individual exposure levels but higher numbers of exposed persons. The analysis of these three parameter combinations is sufficient for now. Chapter 8 discusses the uncertain parameter space in depth.

7.1.1 Benchmark scenario: a close contact leading to high-risk exposure

The close contact scenario that I first introduced in the modeling paper [17] serves as a benchmark scenario to compute a high-risk exposure. High-risk exposure occurs when an infectious person and a susceptible person stand less than 1.5 m apart for over 10 min. During the COVID-19 pandemic, such a situation was commonly associated with a high chance of infection for persons without adequate protection through masks or antibodies (see Section 2.2.1). The simulation result for the parameter sets from Table 7.1 is shown in Fig. 7.1. In the remainder of this work, I use the reference values for exposure of $E_{avg} \approx 9{\cdot}10^2$, $E_{conc} \approx 2{\cdot}10^4$, and $E_{spread} \approx 2.2{\cdot}10^2$ particles to quantitatively distinguish critical from less critical situations.

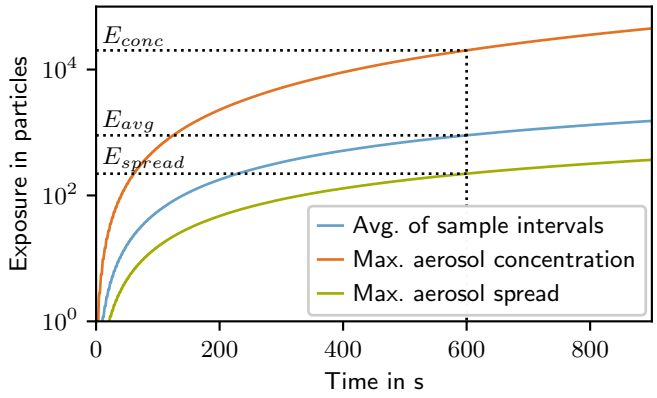


Figure 7.1: The close contact scenario yields different reference exposure levels E_{avg}, E_{conc}, and E_{spread} for the parameter sets from Table 7.1 *avg. of sample intervals*, *max. aerosol concentration*, and *max. aerosol spread*, respectively. These exposure levels vary in the order of magnitudes. Note: own figure adapted from [213, p. 5].

Table 7.1: The parameter sets represent special cases within the bounds of the parameter ranges defined in Table 5.1. Parameter set 1) corresponds to the average of the parameter intervals, while the parameter combinations 2) and 3) lead to the maximum aerosol concentration in an aerosol cloud and the maximum spatial spread, respectively.

Parameter	**Parameter sets**			**Unit**
	1) *Avg. of sample intervals*	2) *Max. aerosol concentration*	3) *Max. aerosol spread*	
Half-life	$5.3 \cdot 10^3$	10^4	10^4	s
Initial radius	1.0	0.5	1.5	m
Dispersion factor	$3 \cdot 10^{-3}$	0	$6 \cdot 10^{-3}$	$\mathrm{m\,s^{-1}}$

7.1.2 Restaurant scenario: COVID-19 outbreak in a restaurant setting

The first validation scenario represents a relatively small superspreading event in a restaurant in Guangzhou, China, in January 2020. Lu et al. [73] report that three groups of people with (A) five persons, including the index case, (B) three persons, and (C) two persons had lunch at neighboring tables. Group A stayed with group B in the same room for 53 min and with group C for 73 min. All members of the three groups tested positive for SARS-CoV-2 in the days after the event. The index patient infected at least one person from the other two groups during the restaurant stay, while the secondary cases from group A could also have been infected outside the restaurant. It cannot be excluded that one secondary case in group B and in group C started a chain of infections among the groups in the subsequent days. The setting with all three groups is shown in Fig. 7.2. Lu et al. [73] mention that air conditioning might have affected the transmission dynamics. However, possible airflow patterns and boundary conditions are not quantified and remain speculative. Keeping in mind that airflow is ignored, I argue that at least one of the members of groups A, B, and C should be considerably exposed in my simulation. If none of the parameter sets produces an outcome matching this condition, the model should be rejected for the respective setting.

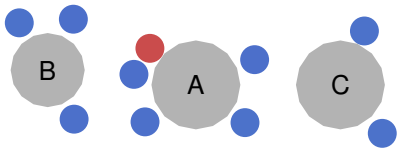


Figure 7.2: In the restaurant scenario, members of three families A, B, and C occupy adjacent tables (gray). The seating chart follows the description in [73]. Secondary cases are depicted in blue, the index case is depicted in red. Note: own figure adapted from [17, p. 14].

Fig. 7.3 summarizes the simulation results. The parameter sets *avg. of sample intervals* and *max. aerosol spread* yield high-risk exposure in all members of groups A and B. This is a first argument in favor of model validity. The members of group C experience lower exposure because they sit further away from the infectious person. Nonetheless, parameter set *max. aerosol spread* causes one and almost a second high-risk exposure in group C. For the *avg. of sample intervals*, the highest exposure in group C is roughly a third of the high-risk exposure. It is not implausible that an infection starts even if the exposure remains at this level. As expected, the parameter set *max. aerosol concentration* leads to far higher exposure than the other two parameter combinations but also in much fewer persons, more precisely only in one member of group A. All other persons are not exposed. Consequently, this extreme case does not reflect the situation adequately. Overall, the model roughly reproduces the COVID-19 outbreak among the three families, except for the *max. aerosol concentration* parameter combination.

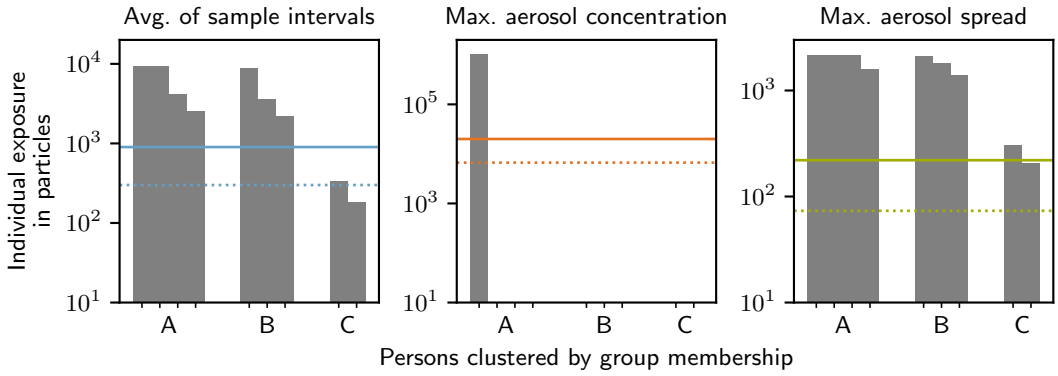


Figure 7.3: The bar charts summarize the individuals' exposure occurring in the restaurant scenario for the three parameter sets from Table 7.1. The (dotted) horizontal lines correspond to (33 % of) the reference exposure observed in the close contact scenario (see Fig. 7.1).

7.1.3 Choir scenario: superspreading event during a choir rehearsal

The second validation scenario is more dynamic and involves more secondary infections than the restaurant scenario. I consider a superspreading event that occurred during a choir rehearsal in Skagit County, Washington, in March 2020. Hamner et al. [129] found that 33 attendees of that choir practice tested positive for SARS-CoV-2 and additional 20 attendees became ill. However, it is not sure that all 20 ill persons were infected with COVID-19 because they were not tested. One person first classified as a probable case tested negative for SARS-CoV-2. In sum, the report lists one index case, 32 confirmed secondary cases, and 20 probable cases.

According to the schedule stated by Miller et al. [71], the singers practiced in groups in two rooms, a large one and a smaller one:

P1) All participants practiced in the large room for 45 min. Several seats between the singers remained empty.

P2) The singers split into two groups of about equal size for 45 min. One group moved to the small room. The other group, including the index case, remained in the large room, occupying approximately half of the initially used space.

P3) The split session was followed by a break of 10 min, where the choir members mingled. Some, including the infectious person, used the restroom.

P4) The final session with all the singers in the large room took another 50 min. The seating chart was the same as during the initial session P1).

In the simulation, I slightly adapt the disclosed schedule to reduce the effort of setting up and evaluating the scenario. The simulated practice session P4) takes 45 min instead of 50 min. I also neglect the relatively short phases of distributing and stacking up chairs before and after the practice, so the chairs are already present before the rehearsal starts.

The chairs are arranged as described in [71], while I design the floor plan to fit the purpose of a choir practice. The actual floor plan is not available. The scenario is depicted in Fig. 7.4.

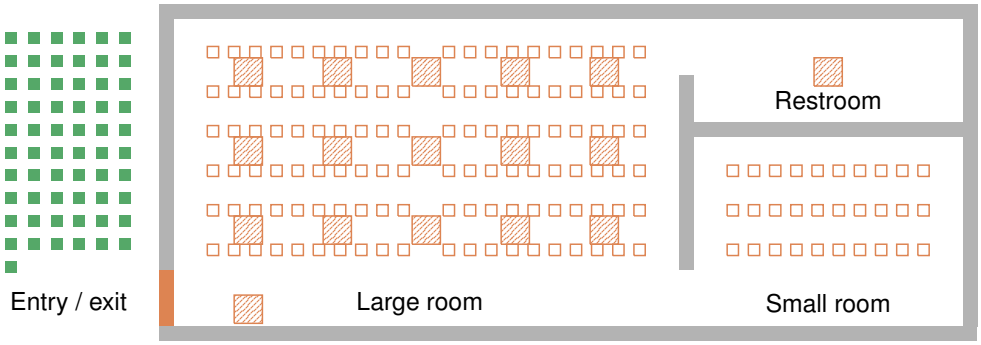
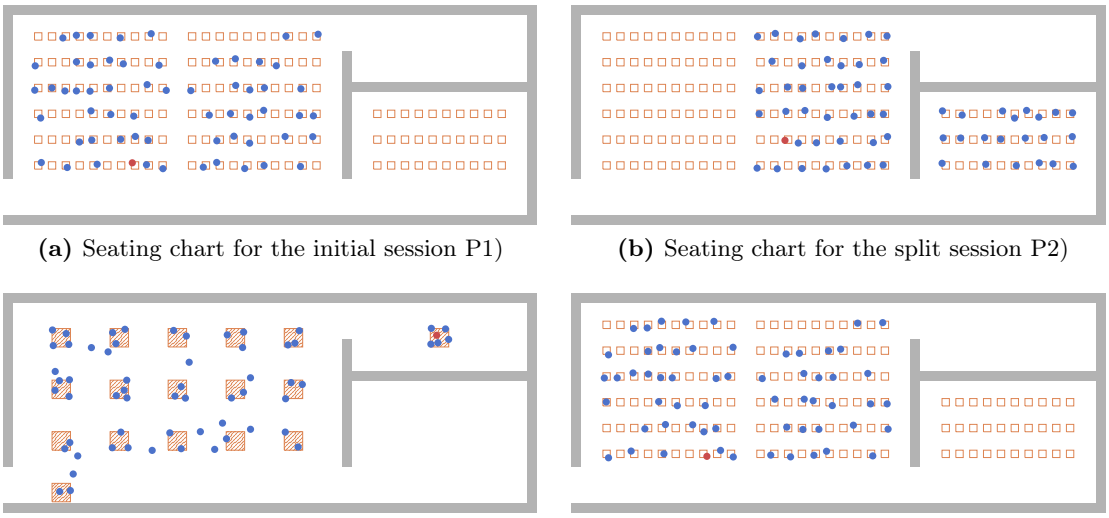


Figure 7.4: The simulation setup for the choir practice is based on a description in [71]. In the large room, 120 chairs (small squares) are arranged in two stages. Within one stage, the center-to-center distance between the chairs is 1.4 m and 0.75 m to the front/back and sides, respectively. The arrangement in the small room is similar. Large, hatched squares represent spots where smaller groups of virtual persons mingle randomly during the break. Virtual persons enter the simulation through the sources (green squares) and leave it through the final target (orange rectangle). The overall size of the scenario is approximately 28 m × 12 m. Note: own figure adapted from [17, p. 15].

To protect the participants' privacy, neither Hamner et al. [129] nor Miller et al. [71] provide a seating chart. This information would be important because the singers' positions affect the individuals' exposure. To address this uncertainty, I repeat the simulation $n = 100$ times with altered seating charts. This means that the sequence of chairs a person occupies changes randomly between two simulations, but the positions of the chairs stay the same. These sequences take into account that the index case stays in the large room during the split session and uses the restroom during the break. Furthermore, the occupancy levels of the blocks in the large room are relatively balanced, as stated in [71]. Fig. 7.5 visualizes several time steps of one of the simulations.

Similarly to the analysis of the restaurant scenario, I conduct numerical experiments for the three parameter sets from Table 7.1 and compare the outcome to the empirical measurements. To that end, I run the set of $n = 100$ repeated simulations for each parameter combination and summarize the output in the histograms in Fig. 7.6.

Most simulations for the parameter sets *avg. of sample intervals* and *max. aerosol spread* yield the same order of magnitude of high-risk exposure levels as confirmed secondary cases. If we consider a lower exposure level of approximately a third of the high-risk exposure, the distribution slightly shifts toward the right, matching the number of confirmed and probable cases even better. It is plausible that the simulation yields fewer persons with high-risk exposure than the actual number of confirmed cases. Three effects mentioned in Chapter 2 are neglected in the model. Firstly, singing typically results in higher pathogen emission rates, whereas the reference scenario of a close contact



(a) Seating chart for the initial session P1)

(b) Seating chart for the split session P2)

(c) Persons mingling during the break P3)

(d) Seating chart for the final session P4)

Figure 7.5: The figures show a fictitious seating chart the virtual persons (susceptible – blue, infectious – red) follow during the four phases of the choir rehearsal. The scenario is the same as in Fig. 7.4, but only the relevant details are displayed. Note: own figures created with Vadere [220].

relates to breathing. Secondly, short-range transmission during the break is not included in the model. Thirdly, the model considers a population with an average susceptibility, which might not apply to the choir members who attended the rehearsal. There is no information about the singers' health condition. However, a median age of 69 years among the attendees [129] suggests that high-risk exposure could occur already at lower levels because older people are generally less immune (see Section 2.1.1). As expected, the parameter set *max. aerosol concentration* leads to fewer individuals reaching high-risk exposure, but the overall number of exposed persons does not compare well with the number of confirmed cases. Even assuming lower exposure levels does not change this picture. Therefore, this parameter set represents an extreme case, and input values beyond this mark should not be used for standard situations. In general, the other two parameter sets return plausible results and, therefore, support the validity of my model.

7.2 Predicting exposure risks for selected everyday situations

In this section, I select specific everyday situations and predict the risk of pathogen transmission. These analyses largely correspond to a part of the results section in my parameter study [213], including the simulation data supplemented in Appendix E (DS5). Some of the simulated scenarios include measures against SARS-CoV-2 transmission, but it is not my primary goal to determine which measure would have been the most suitable to stop the COVID-19 pandemic. I demonstrate the capabilities of my simulation tool

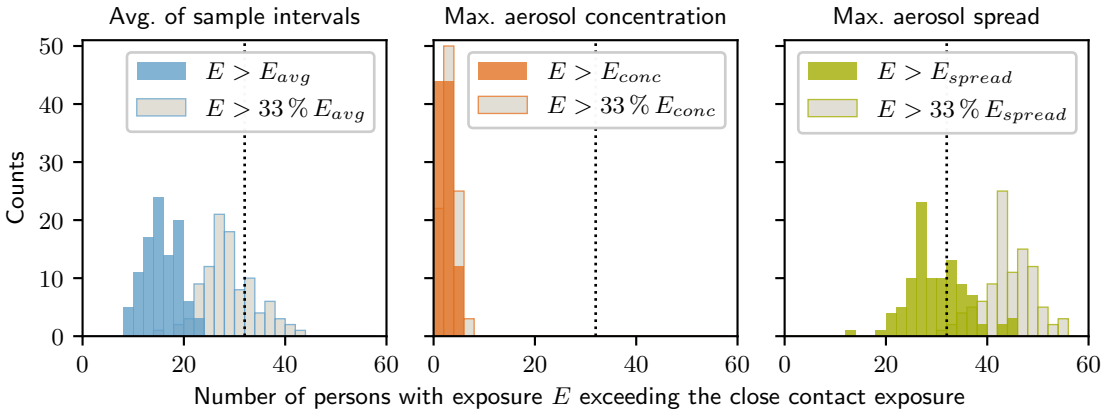


Figure 7.6: The histograms summarize exposure occurring in repeated simulations of the choir scenario for the three parameter sets from Table 7.1. Each parameter set is simulated $n = 100$ times with the virtual persons occupying seats in a randomized sequence. The counts indicate the number of simulations with the respective number of persons exceeding the high-risk reference exposure (colored) or 33 % of the reference exposure (gray) from the close contact scenario. The vertical dotted line corresponds to the 32 empirically confirmed secondary cases.

and how it should be used to identify exposure risks systematically with the aid of quantitative comparisons.

To identify relevant use cases for the simulation studies, I apply several criteria an everyday situation should satisfy. Above all, I concentrate on indoor environments with poor ventilation. Otherwise, high pathogen concentrations are unlikely to occur. For the same reason, the time a person spends in the considered environment should cover at least 10 min. I assume this is the minimum time to elapse until problematic pathogen concentrations can be exhaled or inhaled through aerosols. Furthermore, the analyzed scenarios involve just the right amount of crowd movement. The crowd is neither stationary nor too agile because motionless scenarios do not require a crowd model to solve the problem, whereas highly dynamic scenarios typically lead to comparatively low exposure risks. Instead, the virtual persons change their positions from time to time as it happens, for example, when lining up in a queue. Here, I benefit from the optimal steps model because it can capture human locomotion in various environments even in the case of atypical movement patterns, such as physical distancing. Transmission while queuing in front of a supermarket checkout or during the check-in and boarding process at an airport are realistic settings. Majra et al. [75] find that two reported superspreading events related to air travel primarily occurred before the flight. This gives me a reason to examine such scenarios more closely. In addition, I analyze a relatively dynamic scenario where pedestrians move along an indoor corridor, for example, through the concourse of an airport from one gate to another.

7.2.1 Queuing scenarios: crowd passing a service unit

Motivated by the criteria described above, I now define specific situations in which several people are waiting to be served one after the other at a service station. The settings could, for example, resemble the check-in area at an airport or walk-in centers of transport, health, or other public services. While the floor plan of the building and the queue formation vary, the scenarios all share that the waiting crowd consists of 21 persons. One person is infectious and steps in after the first ten persons have entered the simulation. The inter-arrival time is set to 10 s. The time required to attend to one person at the service unit, referred to as service time, is 90 s. Hence, the total simulation time until the last person leaves the environment is slightly more than half an hour. The assumptions about the arrival and service process are reasonable as long as I am only interested in average values for a queue that has not been further specified. If I aimed to determine the transmission risk in the waiting room of the walk-in clinic in 123 Main Street, Anytown, USA, on a typical Monday morning, it would be necessary to examine the parameters of the queuing system and their distributions in detail. As this is not the case, I continue with exemplary yet simplifying assumptions. This also avoids additional stochastic effects in the model and, thus, reduces the number of model evaluations required to reach convergence of the expected transmission risk.

7.2.1.1 Self-organized queue

Two basic queuing scenarios are depicted in Fig. 7.7. The individuals' behavior solely depends on accepted social norms, while specific queue management strategies are not implemented. The different queue formations reflect cultural variation. For example, Sweden and the United States are categorized as high-queuing countries, whereas lining up tends to be less common in Mainland China or Switzerland. However, such generalizations should be made with caution [299]. Being aware that my terminology may be sociologically imprecise, I describe the queue in Fig. 7.7a as cooperative because the crowd commonly agrees upon the first-come-first-served principle. In contrast, cutting the line is allowed in Fig. 7.7b. I call this a competitive queue because such behavior is linked to the principle of the right of the stronger [300]. Using Vadere's optimal steps model, the competitive behavior corresponds to the default simulation settings. The cooperative queue can be achieved with a dynamic navigation field that takes into account high densities in crowds in such a way that virtual people move toward crowded areas [301].

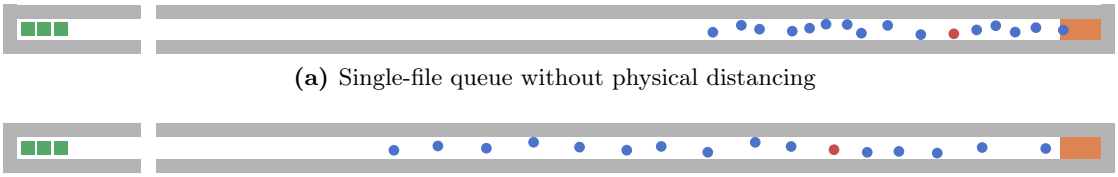
(a) Cooperative self-organized queue

(b) Competitive self-organized queue

Figure 7.7: The snapshots of two simulations demonstrate how the waiting crowd (susceptible – blue, infectious – red) forms a self-organized queue in front of a service unit (orange rectangle). Queue management strategies are not explicitly applied. Note: own figures adapted from [213, p. 3].

7.2.1.2 Single-file queue

A simple strategy to manage the queue according to the first-come-first-served principle is to install barriers. This usually prevents cutting and, in the case of a narrow aisle, results in a single-file queue (see Fig. 7.8). Typical use cases for queue barrier systems include reception or check-in areas, exhibitions, and various other events. They often wind across the available space. However, I simulate a straight line to exclude an additional parameter, the breadth of the queuing system, that could affect the individuals' exposure. Furthermore, it is easier for the locomotion models in Vadere to capture physical distancing in a straight line. The distancing conditions in Fig. 7.8a are relatively normal, while the simulation visualized in Fig. 7.8b accounts for physical distancing of about 1.5 m.



(a) Single-file queue without physical distancing

(b) Single-file queue with approximately 1.5 m physical distancing

Figure 7.8: The snapshots of two simulations visualize several people (susceptible – blue, infectious – red) waiting in single-file queues in front of a service unit (orange rectangle). Note: own figures adapted from [213, p. 3].

I determine the physical distance through a parameter that incorporates the theory of personal space into the optimal steps model (`pedPotentialPersonalSpaceWidth`) [244]. The parameter does not directly translate into a physical distance but alters the underlying utility. In my scenarios, the default value of 0.5 (dimensionless) corresponds to the standard conditions (Fig. 7.8a) and 1.5 to physical distancing (Fig. 7.8b). Fig. 7.9 shows the respective center-to-center distances between two neighboring individuals and the resulting time average. I repeat the simulation $n = 10$ times for the same parameter set but varying sequences of random numbers used in the simulation, which are determined

by the so-called simulation seed. The results confirm that the seed has little influence on the physical distance. The time average of the repeated simulations is a good approximation for either distancing scenario. The exact average values are 0.87 m for normal circumstances and 1.51 m for physical distancing conditions.

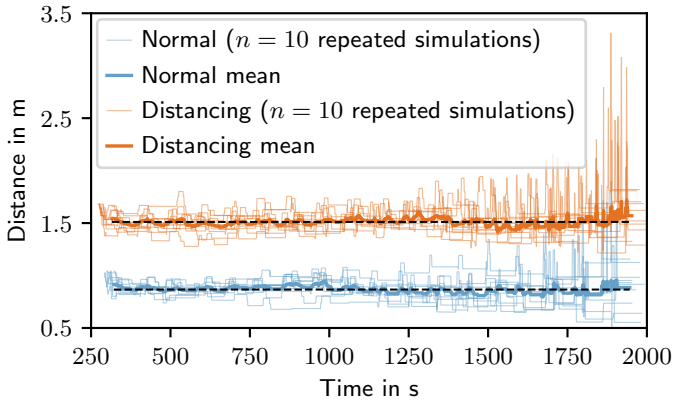


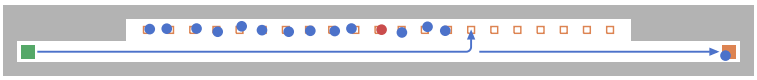
Figure 7.9: The diagram shows inter-individual distances for two conditions in the single-file queue. A thin line represents the average distance between all neighboring persons waiting in the queue. Peaks in these time series are not artifacts but occur every time a person passes the service unit and the other persons successively move up. The more persons are still in the queue, the longer-lasting and lower the peaks. Thick lines build the arithmetic mean of the $n = 10$ repeated simulations (thin lines). Dashed black lines depict the respective time average. Note: own figure adapted from [213, p. 4].

7.2.1.3 Seating scenario with fixed queuing positions

My first simulation study [17] reveals that a single-file queue can cause critical exposure levels in the persons following the infectious person. Introducing a waitlist or ticket system helps to reduce the exposure because it prevents queuing persons from waiting in the trace of aerosol clouds exhaled by the index case. To that end, each person is allocated a fixed position in the waiting room upon arrival and stays there until the previous persons have passed the service unit. Typical examples of such a setting are waiting areas at an airport or in healthcare facilities.

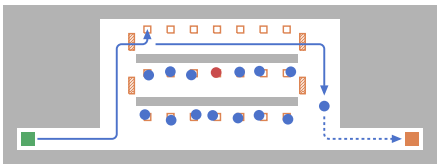
In the parameter study in [213], I analyze the effect of such queue management strategies on exposure risks. I define two scenarios with seats arranged in one row and two scenarios with three rows (block layout) (see Fig. 7.10). For comparability, either layout is simulated with lateral center-to-center spacing between the chairs of 0.85 m and 1.50 m. This corresponds roughly to the normal and physical distancing conditions in the single-file queue, respectively. In reality, many waiting rooms use the available area to full capacity. Consequently, there is often little space between the chairs. Furthermore, people tend to keep a comfortable distance from occupied seats if they are allowed to decide where they would like to wait [302]. However, my setting is particularly plausible

for times of the COVID-19-19 pandemic. At the same time, I would like to emphasize that these scenarios do not represent an actual situation. Their primary purpose is to demonstrate Vadere's capability of reenacting specific everyday situations and movement patterns.

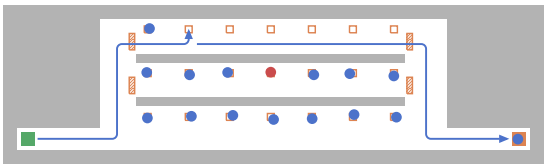


(a) Line layout

(b) Line layout with physical distancing



(c) Block layout



(d) Block layout with physical distancing

Figure 7.10: The snapshots of four simulations visualize several people (susceptible – blue, infectious – red) waiting in a waiting room with seating (orange squares) until it is their turn to be served at the service unit (filled orange square). Intermediate targets (hatched orange rectangles) ensure that the individuals take the right path toward their seats. The seats are occupied (row-wise) from right to left. In (c) and (d), the first arriving person approaches the chair in the upper right corner. Note: own figures adapted from [213, p. 4].

7.2.2 Corridor scenario: counter flow in an indoor corridor

Lastly, I model a situation with crowd dynamics that differs from those in the queuing scenarios. The setting is a long indoor walkway, such as a concourse connecting an airport terminal building with a departure gate. Travelers walk in both directions. A detail of such a counterflow scenario is depicted in Fig. 7.11. Since the pedestrians are constantly in motion, exhaled aerosol clouds cannot accumulate and inhaled pathogen doses should be low. Consequently, I expect minor exposure levels even if a susceptible person directly follows the infectious person.

Figure 7.11: The snapshot of a simulation shows several people (susceptible – blue, infectious – red) walking along an indoor corridor in both directions. The corridor has a length of 150 m. Note: own figure adapted from [213, p. 4].

7.2.3 Simulation results

This section summarizes the model output for the queuing scenarios and the corridor scenarios. Similarly, as for the model validation, I run simulations for the parameter combinations *avg. of sample intervals*, *max. aerosol concentration*, and *max. aerosol spread* defined in Table 7.1. Again, I expect the highest exposure levels for the *max. aerosol concentration* parameter set, whereas the conditions for the *max. aerosol spread* should lead to the highest number of exposed persons. A comparison of the individuals' exposure with the benchmark scenario of a close contact helps me to determine whether critical exposure occurs.

Fig. 7.12 visualizes the individuals' exposure observed in the queuing scenarios. The collection of subcharts presents the results column-wise for the scenarios and row-wise for the parameter sets. For example, the upper left chart, which displays the results for the self-organized cooperative queue simulated with the parameter set *avg. of sample intervals*, unveils that seven individuals exceed the reference exposure of the close contact scenario. If the persons wait in single-file or at fixed positions determined by seating, only two persons experience a high-risk exposure. In the case of physical distancing, the number of persons with high-risk exposure is even zero. The results for the other two parameter sets *max. aerosol concentration* and *max. aerosol spread* show a similar trend. Consequently, it appears reasonable to rank the scenarios according to the number persons with high-risk exposure or the total exposure. Counting only the number of persons with high-risk exposure could give the impression that the physical distancing scenarios of a single-file queue or seating arranged in a line are least critical, whereas the self-organized queues are the most risky situations. However, Fig. 7.12 only provides information about three of many possible outcomes. An average ranking could look different but remains unknown. Therefore, it is vital to consider the uncertainty of the whole parameter space. Chapter 8 delivers a thorough analysis and discussion of the simulation outcome.

In contrast to the queuing scenarios, the corridor scenario leads to much lower exposure for all three parameter combinations. The maximum exposure reached in the simulation remains well below 1 % of the close contact exposure.[1] The primary cause

[1] The maximum possible exposure for the most critical parameter set *max. aerosol concentration* can also be estimated by the equation $\frac{l}{T\,v}\,R_a\,C_{P0} \leq 10^2\,\text{particles} = 0.5\,\%\,E_{conc}$, with a corridor

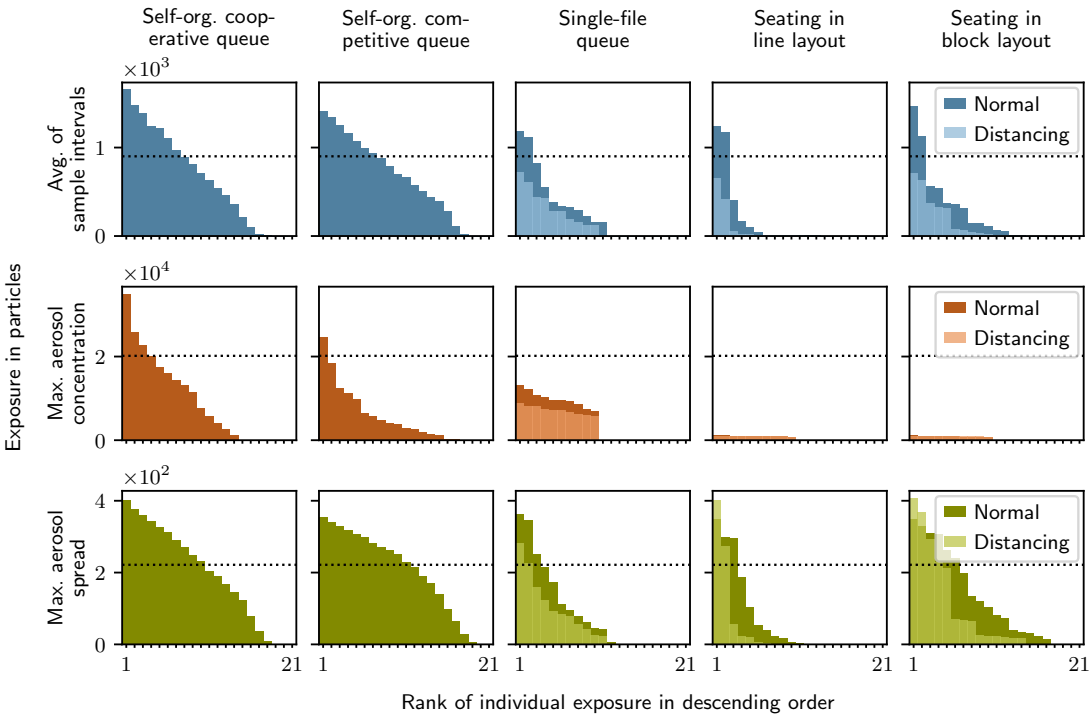


Figure 7.12: The bar charts compare the individuals' exposure obtained for the queuing scenarios defined in Section 7.2.1. Each subdiagram corresponds to the results of one queuing pattern and one parameter combination. The simulation results for the self-organized queue are averaged over $n = 10$ repeated simulations with different seeds (random numbers). Physical distancing is only considered in the single-file queue and seating scenarios. The individuals' exposure is sorted in descending order to increase comparability. The dotted lines indicate a high-risk exposure as observed in the close contact scenario. Note: own figure adapted from [213, p. 6].

for such exposure levels is the marginal pathogen concentration associated with aerosol clouds that do not accumulate. In addition, the virtual persons pass the corridor quickly. Thus, simulations can show comprehensibly that the exposure risk is negligible in such a situation. In the context of the COVID-19 pandemic or future pandemics, this quantitative comparison of scenarios could be a relief for people who feel unsafe in public spaces or help decision-makers who need to devise effective regulations against pathogen transmission. Since the preliminary analysis reveals, as expected, that the corridor scenario is not critical, I scrutinize solely the uncertainty and sensitivity of the queuing scenarios in Chapter 8.

length $l = 150\,\mathrm{m}$, respiratory cycle period $T = 3\,\mathrm{s}$, minimum walking speed $v \geq 0.5\,\mathrm{m\,s^{-1}}$, absorption rate $R_a = 5 \cdot 10^{-4}\,\mathrm{m}^3$ per inhalation, and the initial particle concentration of an aerosol cloud $C_{P0} = \frac{3}{4}\frac{N_{P,\,out}}{\pi\, r_0^3} = \frac{3}{4}\frac{10^3\,\text{particles}}{\pi\, 0.5^3\,\mathrm{m}^3}$. Exponential decay is neglected here because the susceptible person immediately follows the infectious person.

7.3 Summary

This chapter assessed simulations of pathogen transmission via inhalation of airborne aerosol particles in unventilated indoor environments. To that end, I used Vadere's optimal steps model combined with the new feature for transmission via aerosol clouds. In the absence of a biologically plausible dose-response model for SARS-CoV-2, the model returns a person's degree of exposure, which can be interpreted likewise as a theoretical number of inhaled aerosol particles or RNA copies. I simulated a close contact situation to estimate an exposure associated with a high risk of infection. This value served as a reference value for subsequent simulations and helped me to contextualize and interpret respective model outcomes.

My numerical experiments involved three parameter combinations, which I selected from possible ranges defined in Chapter 5. They represent the mean values of the parameter ranges (*avg. of sample intervals*) and two edge cases leading to the maximum aerosol concentration within an aerosol cloud (*max. aerosol concentration*) or its maximum spatial spread (*max. aerosol spread*).

I validated the model for these parameter sets by reenacting two superspreading events and comparing the output with empirical measurements. This comparison showed plausible results for the conditions *avg. of sample intervals* and *max. aerosol spread*, which builds trust in the model. The parameter set *max. aerosol concentration* should be regarded as an unlikely extreme case for the analyzed scenarios.

Encouraged by the model validation, I predicted the exposure risk for pedestrians lining up in a queue or passing an indoor corridor. In summary, the simulations showed that the corridor scenario appears uncritical, while most queuing scenarios involve high-risk exposure. More precisely, the results indicated that the queuing behavior affects the exposure risk. I found considerably higher individual exposure and more exposure levels reaching high-risk in the analyzed self-organized queues than in the queues managed with specific strategies. Accounting for parameter uncertainty is crucial to improve the reliability of these simulations. The next chapter examines the uncertainty in the model input and output in detail.

8 Quantifying uncertainties in predictions of the exposure model

The previous chapter showed that the model output for the degree of exposure can vary in the order of magnitudes depending on the chosen input parameters. Although research on SARS-CoV-2 has progressed, several parameters remain unknown or at least uncertain and diminish the reliability of simulation outcomes. As discussed in Chapter 4, there are mathematical methods to quantify uncertainties. However, uncertainty quantification is not considered as a standard in epidemiological simulations. I aim to bring these fields together and, thus, to increase the reliability and informative value of my previous simulations.

There are several ways to approach uncertainties in simulations. I focus on global sensitivity analysis and forward propagation. It would not make sense to apply inverse methods because they require data that are not available. By running a sensitivity analysis and an uncertainty analysis, I predominantly seek to answer the following questions and their respective implications for reality:

Research questions

How can uncertainties in the simulation be quantified?

- Sensitivity analysis: How can uncertain input parameters be prioritized regarding their impact on the output?
- Uncertainty analysis: How can one quantitatively determine the effect of input uncertainties on individual exposure levels and the number of exposed persons in a scenario?

Appropriately targeting these questions demands a large number of model evaluations. Unlike in Chapter 7, this task exceeds the resources of an ordinary computer if the computing time is to remain below acceptable limits. Consequently, I also describe how I tackle this computational problem.

This chapter is structured as follows: First, I briefly recapitulate the relevant uncertainty quantification methods and explain which tools I use to conduct my numerical experiments. Then, I determine and discuss the model sensitivity and output uncertainty. I covered these analyses in my parameter study [213], yet with fewer details about the tooling. Gerta Köster and Hans-Joachim Bungartz co-authored this publication.

8.1 Methods for sensitivity analysis and forward propagation

I adopt a methodology as schematically shown in Fig. 8.1. Instead of considering merely point estimates, I characterize input uncertainties by probability distributions and prop-

agate them through the model. Putting the variability of the output in relation to the input uncertainty delivers the model's sensitivity. Thus, I distinguish influential from non-influential model parameters. In addition, a statistical analysis of variations in the output improves the reliability of the simulations.

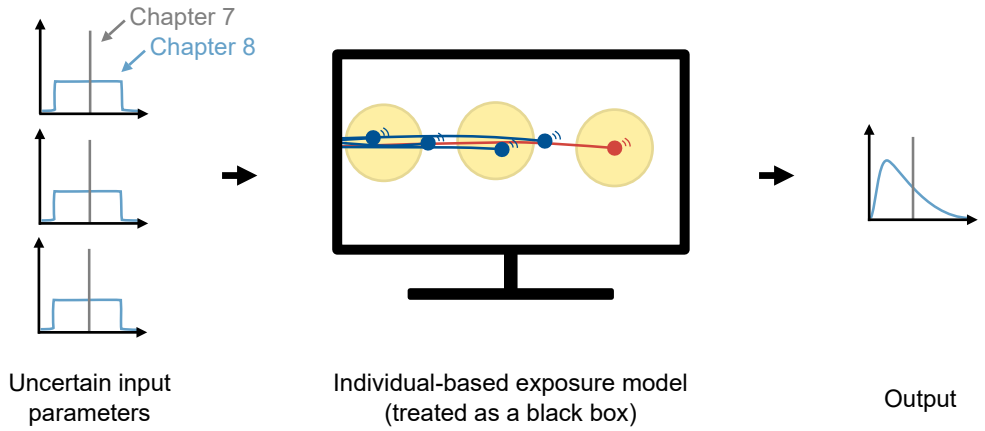


Figure 8.1: The illustration schematically compares the approaches pursued in Chapter 7 and this chapter. While I previously evaluated point estimates, I now consider the whole parameter space. This means I feed probability distributions into the model to obtain a complete picture of the output uncertainty. The uncertainty quantification methods I use treat the model as a black box. Note: own figure adapted from [213, p. 2].

For sensitivity analysis, I employ a global method to capture the impact of parameters completely. More precisely, I choose the Sobol' method (see Section 4.2.1) because it is applicable to nonlinear nonmonotonic models without interfering with the program code of the model. It returns total sensitivity indices, S_{Ti}, which quantify the sensitivity of the model output with respect to the uncertainty of the i-th parameter and interaction effects between the i-th parameter and other parameter uncertainties. First order sensitivity indices, S_i, describe the isolated impact of parameter i. Since my numerical experiments yield similar first order and total indices, I focus on the total sensitivity. Another benefit of the Sobol' method is that several toolboxes already offer ready-to-use implementations. I use SALib [271, 272], version 1.4.5. Section 4.4.1 describes the two relevant functions for generating samples and estimating sensitivity indices.

My uncertainty analysis comprises (pseudo-)random Monte Carlo simulations and a statistical summary of the output. Alternative approaches, such as perturbation methods or spectral representations, are usually less computationally demanding. However, they lead to model form uncertainties and, therefore, methodological constraints [258, pp. 187–8]. I generate (pseudo-)random Monte Carlo samples with the Python library chaospy [270], version 4.3.10. Section 4.4.2 provides details about the sampling function.

8.2 Materials: tools, infrastructure, and model settings

Both sensitivity analysis and uncertainty analysis constitute a computationally expensive problem. The Python toolbox suqc,[1] initially developed by Lehmberg [303], helps me reduce the computing time through parallelization. Thus, I can exploit multiple cores for better performance. I execute the simulations described in the following sections on a virtual machine hosted on a Linux server with two processors (Intel® Xeon® Silver Processor 4114, equipped with 10 cores, 20 threads, 2.2 GHz processor frequency, x86-64 architecture) providing 40 threads in total and a random-access memory (RAM) of 57.5 GiB. The respective computing times for a numerical experiment range from several minutes to about two days.

My numerical experiments are based on Vadere 2.2. I connect Vadere and the packages sucq, SALib, and chaospy through a new Python package, which I name vadere-infection-model-uq (vimuq).[2] This repository provides a structure for conducting numerical experiments tailored to the infection module in Vadere. In addition, it contains all simulation studies, that is, the definitions required to run a sensitivity or uncertainty analysis for a particular scenario. The simulation data is stored separately and available in Appendix E (DS6). Thus, others can access and reproduce the numerical experiments. The workflow for conducting a new experiment consists of two separate steps: generating and evaluating samples and then analyzing the model output. Further details are given in the repository, including instructions about how to install the vimuq project.

To avoid errors in the code, I inspect the important parts such that unit tests cover 63 % of the lines. Complementary integration tests ensure that a whole numerical experiment runs without throwing errors. Furthermore, I run my implementation for a frequently employed test case called the Ishigami function [304]. The function is nonlinear and nonmonotonic. I replace Vadere temporarily with this function and calculate sensitivity indices to verify that my framework returns the expected results. The continuous integration pipeline of the vimuq project executes the above routines with every update of the GitLab repository. Thus, I maintain the level of code quality throughout the development process.

I scrutinize the model input and output using the above set of methods and tools. The model itself is treated as a black box. This means, the uncertainty quantification methods only know about the input and outputs, but they do not require access to the underlying implementations or algorithms. The models under investigation are the queuing scenarios introduced in Section 7.2.1. More precisely, I reconsider the self-organized competitive queue, the single-file queue, and the seating scenarios with fixed queuing positions. I only leave out the cooperative queue, which is much the same as the competitive queue.

The model input is linked to the same parameters as in the previous chapter. I still regard the parameters *half-life*, *initial radius*, and *dispersion factor* as uncertain inputs, but they are now tied together mathematically through a joint probability distribution. I assume all three parameters are independent and uniformly distributed within their

[1] www.gitlab.lrz.de/vadere/suq-controller, commit 3b0988ba, accessed on April 2, 2024
[2] www.gitlab.lrz.de/vadere/infection-model-uq, accessed on April 2, 2024

uncertain ranges. The uniform distribution is an initial guess, as there is not enough trustworthy data available to suggest other shapes of distributions. This choice should be adapted when there is sufficient evidence about the true distribution of the parameters. As a result, I obtain the parameter space listed in Table 8.1.

Table 8.1: My uncertainty quantification experiments analyze the impact of the listed parameter uncertainties. The uncertainties are represented by uniform distributions with bounds corresponding to the parameter ranges defined in Table 5.1. I use the log-uniform distribution of the *half-life* for one numerical experiment to confirm that it yields a similar sensitivity as the uniform distribution in linear space.

Parameter / random variable X	**Distribution of X**	**Unit**
Half-life	$X \sim \mathcal{U}\left(6 \cdot 10^2,\, 10^4\right)$	s
Half-life (logarithmic space)	$\log_{10}(X) \sim \mathcal{U}\left(\log_{10}\left(6 \cdot 10^2\right),\, \log_{10}\left(10^4\right)\right)$	s
Initial radius	$X \sim \mathcal{U}\left(0.5,\, 1.5\right)$	m
Dispersion factor	$X \sim \mathcal{U}\left(0,\, 6 \cdot 10^{-3}\right)$	$\mathrm{m\,s^{-1}}$

The parameter *half-life* ranges over two orders of magnitude, so a log-uniform distribution could be more appropriate than the uniform distribution in linear space. Therefore, I also try out one numerical experiment with the *half-life* following a base-10 logarithmic uniform distribution. As shown in the next section, the shape of the distribution has almost no influence. The results for the uniform distribution and the log-uniform distribution are similar for the analyzed scenario. I argue that this applies also to the other scenarios because they are closely related. Consequently, I run all other numerical experiments only with the uniform distribution in linear space because it is easier to understand and delivers roughly the same information. In contrast to the *half-life*, using log-uniform distributions for the parameters *initial radius* and *dispersion factor* would be inadequate. The interval bounds of these two parameters are less than a factor of ten apart or include zero, which cannot be converted to logarithmic scales.

The model output remains the same as in the previous chapter: I evaluate the pathogen or aerosol particles each individual takes in. In addition, I analyze the average exposure among those who become exposed, the number of exposed persons (exposure $E > 0$), and the maximum exposure.

8.3 Sobol' sensitivity analysis for factor prioritization

In this section, I present the results of the global sensitivity analysis that I conducted in [213] to quantify the effect of input variability in the model parameters on the variability of the output. Information about parameter importance fosters a better understanding of the model and reveals where one can effectively improve the reliability of simulations. High sensitivity regarding a specific model parameter can indicate how to control

pathogen transmission effectively. In the following, I demonstrate how I achieve reliable results for the sensitivity indices. Then, I discuss and interpret the results.

8.3.1 Simulation set-up for a reliable sensitivity analysis

It is essential to confirm that the estimates for the model sensitivity are correct before one can discuss them in detail. To that end, I check whether the sample size achieves appropriately converged results. Convergence depends on several factors, including the sampling technique. I use Saltelli's extension of the Sobol' sequence (see Section 4.4.1 for details) to generate samples from the 3D (dimension $d = 3$) parameter space. The respective sample points must be chosen carefully to increase performance compared to classical Monte Carlo sampling. Owen [278] proposes that the base sample size N be a power of two and to skip the first $N_{skip} \geq N$ points. Following this recommendation, I find $N = 2^{12}$ to yield accurate results for the single-file queue. I determine this sample size by step-wise doubling $N = 2^k$ for $k = 4, 5, \ldots, 12$ and evaluating the difference in the resulting total and first order sensitivity indices, $\Delta S = |S(N_k) - S(N_{k+1})|$. The absolute difference suffices since the sensitivity indices range in the interval [0, 1]. As k increases, the difference ΔS shrinks and finally drops to a level I deem acceptable (< 0.01 for most quantities of interest). Thus, I assess that $N = 2^{12}$ yields satisfactory convergence for the single-file queue. I assume this estimate also holds for the other scenarios because they are similar. In sum, computing total and first order sensitivity indices requires $N(d + 2) = 20\,480$ sample points for each scenario.

Importantly, these convergence estimates apply to the model's deterministic form but not generally to a stochastic set-up. As mentioned in Chapter 7, simulations with Vadere usually rely on sequences of (pseudo-)random numbers. Each sequence is associated with a so-called seed. Keeping the seed fixed for repeated simulations of the same scenario and parameter set leads to exactly the same simulation result. Otherwise, mainly the spawning process of virtual persons and pathfinding algorithm introduce randomness. This alters the virtual persons' trajectories and, consequently, their exposure in simulations repeated with different seeds. A single simulation represents only one possible realization, and one needs several realizations to get a meaningful average output. This leads to the questions of (1) how to post-process the simulation output of repeated realizations and (2) which number of repetitions, n, is adequate.

Fig. 8.2 visualizes how I post-process the output in three steps: averaging the output, calculating additional quantities of interest from the individual exposure levels delivered by the exposure model, and determining the model sensitivity. Options A – D in Fig. 8.2 indicate in which order these three operations are executed. The results discussed in the next sections are produced as follows. In the case of the competitive queue, I average the sensitivity indices determined for each repetition (option C). This order is necessary to keep data points apart that should not be mixed prior to calculating sensitivities. For example, it avoids averaging exposure levels that belong to different persons, as the crowd does not adhere to the first-come-first-served principle. For all other scenarios, I first average the output and then calculate the sensitivity indices (option A). Considering

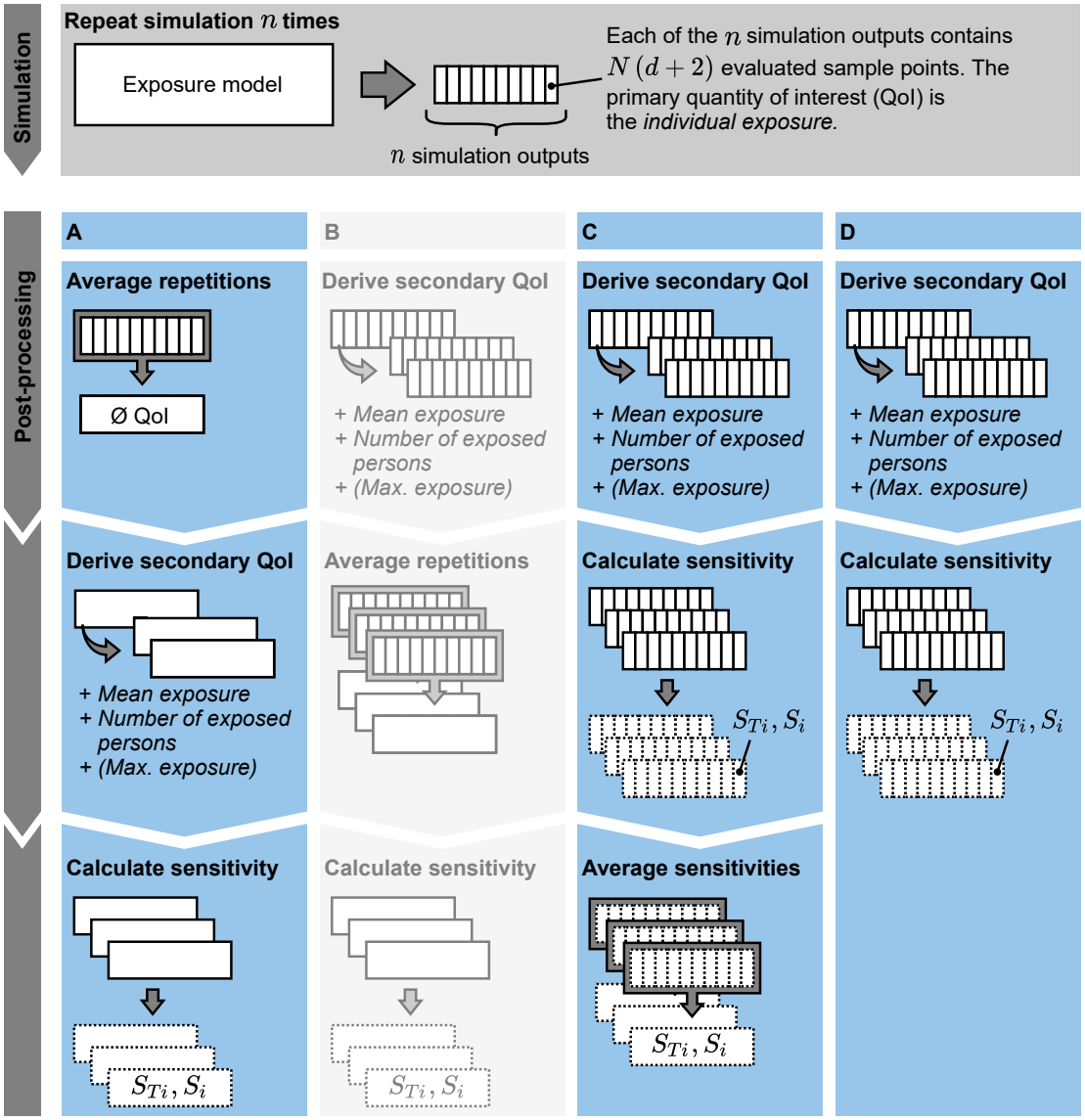
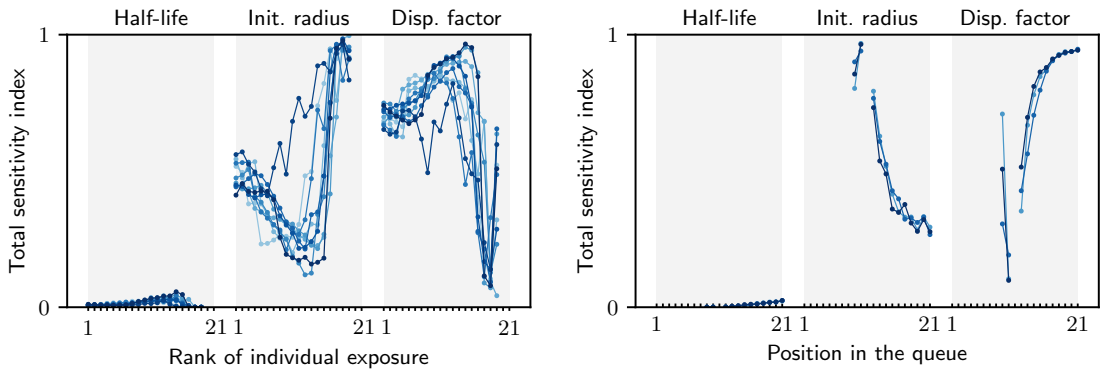


Figure 8.2: The chart shows four methods of post-processing the simulation output. The model returns the individual exposure as a primary quantity of interest. Options A – D indicate in which order I calculate secondary quantities from the primary output, average the results, and compute the total (S_{Ti}) and first order (S_i) sensitivity. I use options A and D for the single-file queue (with $n = 3$ repetitions) and C and D for the competitive queue ($n = 10$). All other scenarios ($n = 1$) are processed according to A, the method with the fewest computational operations.

the sensitivity indices for each repetition separately (option D) is only relevant in the next paragraph, where I confirm that my choice of the number of repetitions, n, is adequate.

A conclusive answer to question (2), relating to the number of required repetitions, would require considerable computational effort. I adopt a simplified approach. I argue, the more confined the virtual persons' movement is, the less influential the stochastic effects of the pathfinding algorithm. Consequently, the queuing scenarios with queue management are likely to reproduce similar individual exposure levels independent of the seed. In contrast, free movement in the self-organized competitive queue can lead to significantly different exposure. Therefore, I make an initial guess and check the differences occurring in $n = 10$ repeated simulations of the competitive queue. Fig. 8.3a shows the total sensitivity indices calculated for each repetition (option D from Fig. 8.2). For now, it is enough to understand that the curves are roughly the same, so the repeated simulations lead to comparable sensitivities. An interpretation of the sensitivity indices follows in the next section. The similarity between the curves supports my argument that, for the competitive queue, considering the average of $n = 10$ repetitions should suffice to obtain reliable information about the sensitivity. I apply the same procedure to the single-file queue, yet with only $n = 3$ repetitions because I expect less randomness in the model. Indeed, Fig. 8.3b demonstrates that even a single simulation should yield reliable sensitivities since the three curves almost overlap. The other scenarios involve even less randomness because of the fixed waiting positions. Therefore, I argue that analyzing only one simulation ($n = 1$) in these cases is reasonable. Note that Fig. 8.3 displays only the total sensitivity indices for the individuals' exposure in the queue. The other quantities of interest also produce similar results for each repetition. This finding holds likewise for the first order sensitivity indices.



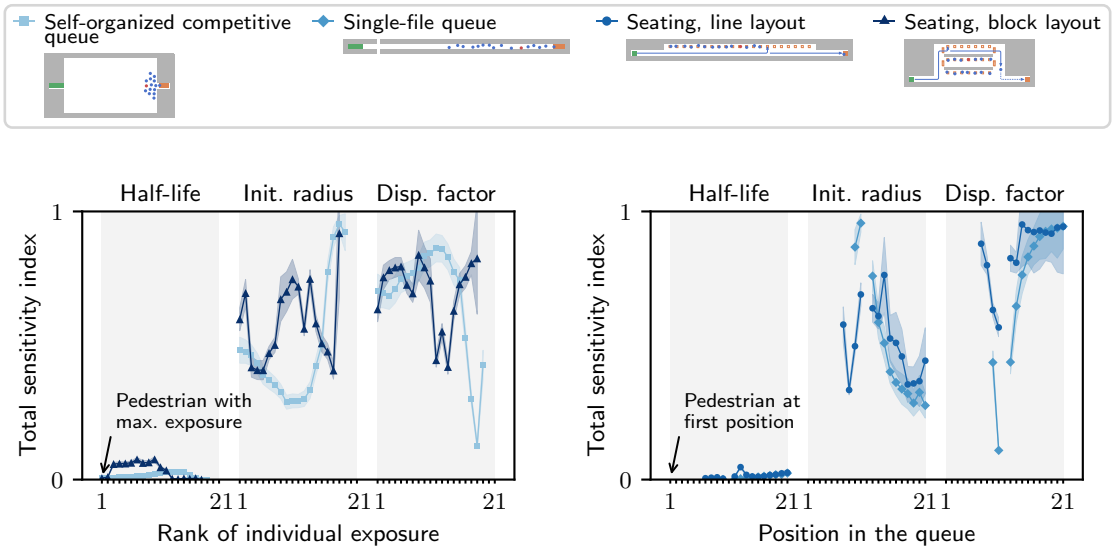
(a) Competitive queue ($n = 10$ repetitions) **(b)** Single-file queue ($n = 3$ repetitions)

Figure 8.3: The charts compare the results of sensitivity analyses repeated with different seeds. They demonstrate that the model stochasticity affects the sensitivity indices only slightly. Each curve represents one repetition (shades of blue). The displayed data set stands exemplary for all quantities of interest and total or first order sensitivity indices. Here, it shows the total sensitivity index calculated for the individuals' exposure and each parameter. Note: own figures adapted from [213, p. 12].

8.3.2 Identifying influential and non-influential parameters

Having ensured an appropriate set-up for the sensitivity analysis, I now present and discuss the resulting sensitivity indices. The aim is to quantify the impact of the three uncertain input parameters on quantities of interest and, thus, to distinguish influential from non-influential parameters. The impact depends on the quantity of interest. This section reviews four quantities of interest separately, starting with the sensitivity of the individuals' exposure. After that, I analyze the average exposure among exposed persons, the number of exposed persons, and the maximum exposure in a scenario. To further reduce the complexity of the numerical experiments, I concentrate on the queuing scenarios without physical distancing and treat physical distancing in a separate section. Moreover, I primarily discuss the results of the total sensitivity. A total sensitivity index describes the impact of an uncertain parameter and its interactions with other uncertain parameters on the model output. The first order sensitivity indices, which include only the isolated impact of uncertain parameters, are similar to the total sensitivity in my numerical experiments. I cover them briefly at the end of this section.

The total sensitivity of the individuals' exposure observed in a scenario varies between the virtual persons, as can be concluded from Fig. 8.4. The quantities of interest differ slightly between Fig. 8.4a and Fig. 8.4b. I consider ranked exposure levels in Fig.



(a) Quantity of interest: individual exposure ranked in descending order

(b) Quantity of interest: individual exposure sorted by queuing position in ascending order

Figure 8.4: The figures show the total sensitivity indices of the individuals' exposure concerning the three uncertain parameters *half-life*, *initial radius*, and *dispersion factor*. The sensitivity index of the infectious person is undefined because the respective exposure is always zero. It corresponds to rank 21 in (a) and to position 10 in (b). The bootstrap confidence intervals (shaded areas) are computed from 100 re-samples at a 95 % confidence level. Note: own figures adapted from [213, p. 7].

8.4a and exposure levels sorted by position in the queue in Fig. 8.4b. This distinction accounts for the fact that the scenarios analyzed in Fig. 8.4a do not allow for mapping the sensitivity indices to certain persons or queuing positions. In contrast, the queues described in Fig. 8.4b are organized in a straight line so that each person can be associated with a particular sensitivity index. While the parameter *half-life* is generally of minor importance, the impact of the parameters *initial radius* and *dispersion factor* depends on the position in the queue. The charts indicate that the *initial radius* rises in importance for susceptible persons who follow the infectious person, while the *dispersion factor* declines with increasing distance to the infectious person. Taking a look at each sensitivity index separately makes this connection more apparent.

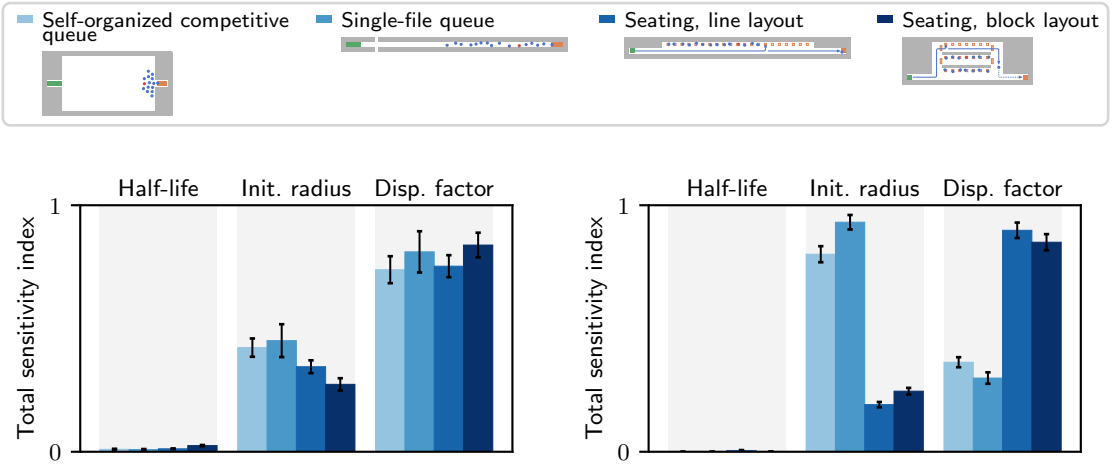
First, I explain the sensitivity of the self-organized competitive queue in Fig. 8.4a. The opposing trend of the parameters *initial radius* and *dispersion factor* results from the following circumstances: Individuals grouping closely around the infectious person experience, on average, the highest exposure. The further away they stand, the lower their exposure. These exposure levels correspond to approximately the first half of the ranks. Persons at distances larger than the *initial radius* of an aerosol cloud only become exposed as the cloud expands or when they walk into clouds. Accordingly, the importance of the *initial radius* declines and the *dispersion factor* becomes more influential. This relationship flips approximately after rank 11, so the impact of the *initial radius* increases, and that of the *dispersion factor* falls steeply. This is mainly a consequence of aerosol clouds dissolving before the persons at the rear of the queue reach the contaminated area. Rank 19, attributed to the person who passes the service unit second, shows an unexpectedly high sensitivity regarding the *dispersion factor*. When averaging the Sobol' index computed from multiple repetitions, it is subject to instabilities. The root cause is that only parameter combinations with large values for the *initial radius* and the *dispersion factor* cause exposure in this position, but many samples lead not to exposure. Consequently, the total variance in the output is smaller than 1, which results in overestimation as the partial variance is normalized by the total variance. Ranks 20 and 21 are not defined because the respective exposure is zero. The individuals corresponding to rank 20 leave the scenario before passing an aerosol cloud, and rank 21 belongs to the infectious person.

Fig. 8.4b shows the sensitivity observed for the single-file queue and the seating scenario with a line layout. Since the 21 pedestrians wait in the same order throughout all simulations, it is valid to calculate a sensitivity index by person or position in the queue. The first six (seating, line layout) to eight (single-file queue) persons are not exposed. Hence, their sensitivity indices are not defined. For the remaining positions, the impact of the *initial radius* and the *dispersion factor* have opposite tendencies in both scenarios. Generally, the influence of the *initial radius* decreases with increasing distance to the infectious person (position 11) to the front and to the back. In contrast, the *dispersion factor* becomes more important the larger the distance. The reasons for this trend are the same as detailed for the competitive queue. Persons near the infectious person undergo relatively high exposure, which is determined by the *initial radius*. The exposure of persons further apart predominantly depends on whether aerosol clouds maintain high

concentrations as they expand slowly or whether the clouds dilute quickly. Looking closer at the sensitivity of the seating scenario reveals several peaks. Most importantly, the peaks at position 14 in the *initial radius* and the *dispersion factor* arise from different places where exposure occurs. Depending on the parameter combination, the respective individual experiences medium to high exposure at the seating position or relatively low exposure at the service unit. This is reflected in an increased sensitivity. The peaking sensitivity of the *initial radius* at position seven in the seating scenario occurs for a similar reason. The *initial radius* is either large enough to cause exposure or too small, so the seventh person remains unexposed.

The total sensitivity indices for the seating scenario with the block layout (see Fig. 8.4a) are difficult to assign to individuals or positions. The data set does not show an explicit tendency, as for the other scenarios. It is included in the chart for the sake of completeness. The average exposure and the number of exposed persons give more insights into this scenario.

The total sensitivity indices for the average exposure among the exposed persons and the number of exposed persons are visualized in Fig. 8.5. These quantities of interest give an overall impression of how risky a situation can be. Note that they both consider a person exposed whose degree of exposure is larger than 0. One could also use another threshold instead of 0, but this would raise the issue of how to choose the threshold and what this threshold should stand for. The reference values calculated for the close contact scenario do not help here because they refer to a specific parameter combination, whereas the sensitivity analysis relies on intervals for the three uncertain parameters. Fig. 8.5a



(a) Quantity of interest: average degree of exposure among the exposed persons

(b) Quantity of interest: number of exposed persons

Figure 8.5: The figures show the total sensitivity indices of the quantity of interest concerning the three uncertain parameters *half-life*, *initial radius*, and *dispersion factor*. The bootstrap confidence intervals (error bars) are computed from 100 re-samples at a 95 % confidence level. Note: own figures adapted from [213, p. 8].

shows that the sensitivity of the average exposure is roughly the same in all four scenarios. The *initial radius* is less important than the *dispersion factor*. Compared to that, the influence of the *initial radius* and the *dispersion factor* on the number of exposed persons varies between the scenarios (see Fig. 8.5b). Interestingly, the *initial radius* is much more influential than the *dispersion factor* for the competitive queue and the single-file queue. In the case of the seating scenarios, the relation between the importance of the *initial radius* and the *dispersion factor* is the opposite. This change in the sensitivities reflects the two distinct queuing patterns of successively moving up versus occupying stationary positions. The continuously moving queues are more responsive to uncertainties in the *initial radius* because the extent of the aerosol clouds determines whether persons ahead of the index case become exposed. Moreover, people at the end of the line regularly move forward and, therefore, reach contaminated areas earlier than in the seating scenarios. In the seating scenarios, persons at the rear of the queue evade exposure if the aerosol clouds vanish before they walk through them. Whether the scenario is cleared from aerosols after a certain time depends primarily on the *dispersion factor*.

The maximum exposure (see Fig. D.1a in Appendix D) is influenced by the *initial radius* and *dispersion factor* to approximately the same extent, regardless of the scenario. The total sensitivity indices of these two parameters are about 0.5, while the *half-life* has no impact.

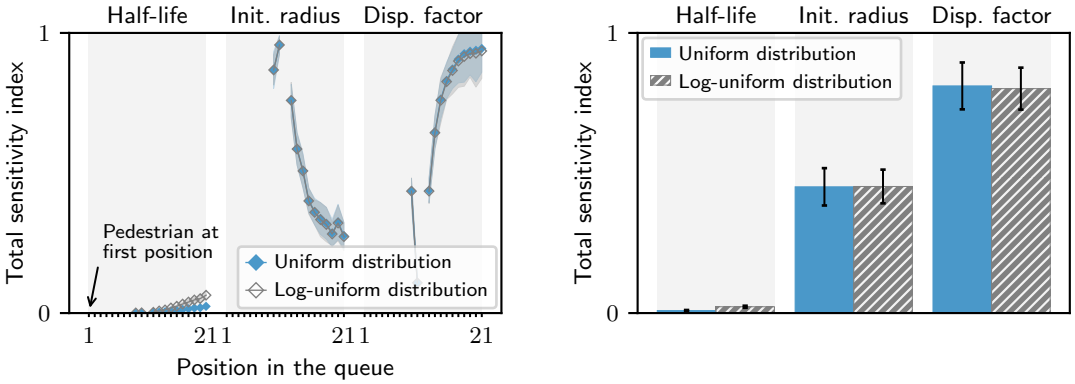
So far, I have only addressed the total sensitivity. The first order sensitivity, which evaluates the isolated influence of a parameter, leads to the same parameter ranking (see Fig. D.1b, D.2, and D.3 in Appendix D). That is, the *half-life* is non-influential, whereas the other two parameters have varying importance depending on the considered quantity of interest. The first order sensitivity indices are slightly smaller than the total sensitivity indices. Hence, interaction effects between the parameters *initial radius* and *dispersion factor* play a role, yet mainly to a limited extent. Interactions with the *half-life* are negligible because the total sensitivities of this parameter are already close to zero.

8.3.3 Sensitivity for a log-uniform distribution of the parameter *half-life*

The previous analyses with a uniform distribution representing the input uncertainty of the *half-life* show that this parameter has almost no impact on the exposure. However, this could change if one assumes a different input distribution. Above all, a log-uniform distribution could be a better choice for the input uncertainty because the *half-life* adopts values across two orders of magnitude (see Table 8.1). A log-uniform distribution of the *half-life* implies that the common logarithm of the *half-life* is distributed uniformly as per $\mathcal{U}\left(\log_{10}\left(6 \cdot 10^2\right), \log_{10}\left(10^4\right)\right)$ in s. As a consequence, the region closer to the lower bound of the uncertain interval is weighted more strongly compared to a uniform distribution in linear space. The other two parameter uncertainties remain unchanged.

I examine the sensitivity of this adapted parameter space for the single-file queue and compare the results to the setting with a uniform distribution of the *half-life*. Fig. 8.6 shows the sensitivities for the individuals' exposure and the average exposure. The results are generally the same. The impact of the *half-life* on the individual exposure levels increases minimally toward the end of the queue, which means the *half-life* becomes

more important the longer a person stays in the scenario. This is reasonable because a shorter *half-life* contributes to faster decay of aerosol concentrations and, thus, reduces the exposure more than it is the case for a uniformly distributed *half-life*. Nonetheless, the effect of right-skewed input samples is negligible for all quantities of interest. It would become significant if the lower bound of the uncertain interval included even shorter times. However, while such a parameter choice may be applicable to other respiratory diseases, it would not be justified for COVID-19 in view of the current knowledge. For this reason, I continue with a uniformly distributed *half-life*, which is easier to comprehend.



(a) Quantity of interest: individual exposure sorted by queuing position in ascending order

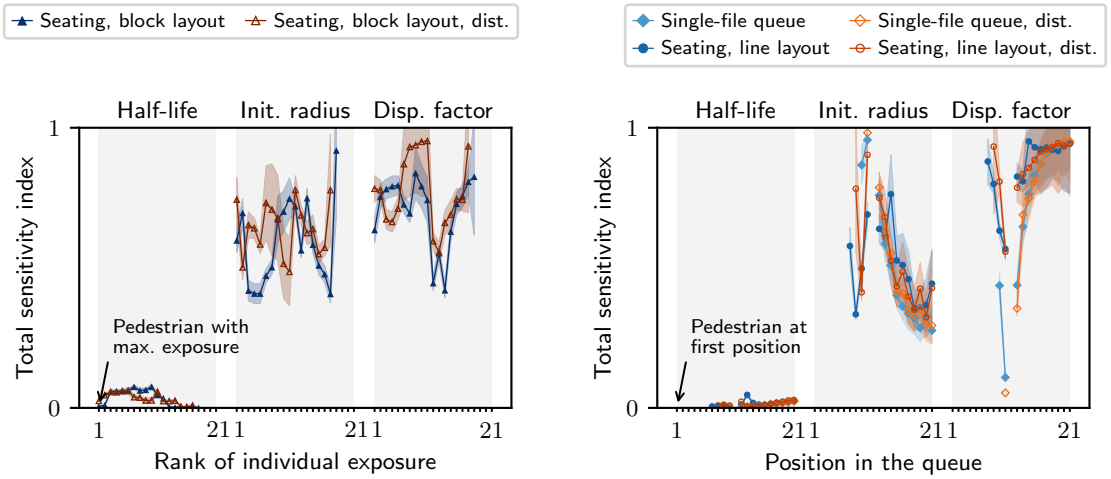
(b) Quantity of interest: average degree of exposure among exposed persons

Figure 8.6: The figures compare the total sensitivity indices for the single-file queue with the parameter *half-life* following a uniform distribution or a log-uniform distribution. The bootstrap confidence intervals (shaded areas / error bars) are computed from 100 re-samples at a 95 % confidence level.

8.3.4 Impact of physical distancing on the sensitivity

The sensitivity analysis also includes scenarios with physical distancing of approximately 1.5 m. Compared to normal conditions, physical distancing prolongs the time between when the infectious person exhales an aerosol cloud and when a susceptible person inhales the particles. Consequently, the aerosol concentration decreases more, and the exposure levels are slightly lower. Interestingly, this produces only small changes in the sensitivity. Fig. 8.7b shows that the trends of the total sensitivity indices essentially remain the same. In the seating scenario with a block layout (Fig. 8.7a), sensitivity indices per rank of exposure seem volatile and different from the normal distancing conditions. However, the general region of the sensitivity is the same. In addition, the sensitivity of the average exposure is similar (see Fig. D.4 in Appendix D). These unchanged sensitivities indicate that physical distancing does not affect the transmission dynamics in these queuing scenarios, which is plausible as long as one only models long-range transmission. If short-

range transmission via larger aerosol particles were included in the model, the sensitivity should be clearly different between scenarios with and without physical distancing.



(a) Quantity of interest: individual exposure ranked in descending order

(b) Quantity of interest: individual exposure sorted by queuing position in ascending order

Figure 8.7: The figures compare the total sensitivity indices for the scenarios without physical distancing (shades of blue) and with physical distancing (shades of orange). The bootstrap confidence intervals (shaded areas) are computed from 100 re-samples at a 95 % confidence level. Note: own figures adapted from [213, p. 13].

8.3.5 Implications of the sensitivity analysis

The sensitivity indices presented above have several implications. I discuss the meaning of sensitivity with respect to modeling and simulation. Then, I approach the sensitivity indices from a practitioner's viewpoint and interpret their consequences for real life.

Most importantly, neither of the distributions of the parameter *half-life* has a relevant impact on the exposure in the scenarios under consideration. For this reason, it makes no difference whether or not the *half-life* is treated as uncertain. One can use a point estimate instead of an interval and, thus, reduce the parameter space and number of model evaluations. I suggest the mean value of either the uniform distribution or the log-uniform distribution.

In contrast to the *half-life*, the uncertainty introduced by the parameters *initial radius* and *dispersion factor* are influential. Both parameters describe distinct aspects of aerosol spread over time. Whether the *initial radius* is more important than the *dispersion factor* or vice versa is linked to the queuing strategy and the observed output. Therefore, both parameters constitute a starting point for effectively improving the aerosol model and reducing related uncertainties. Since the spatiotemporal propagation of aerosol clouds within a room considerably influences the output and the sensitivity indices, it would be reasonable to incorporate airflow into the model. For example, this could be estimated

empirically by learning aerosol dynamics from image data. Alternatively, one could use generalized maps of pathogen concentration around a source derived from CFD simulations as recently proposed by Nicolas and Mendez [209]. Seeing the potential of more detailed approaches on the one hand, I underline the advantage of parsimonious models on the other hand. Too complex modeling techniques prohibit thorough sensitivity analyses.

Finally, choosing a global method for sensitivity analysis appears appropriate since interaction effects between the parameter uncertainties play a role. A local method could not detect these effects.

If we shift the viewpoint from modeling and simulation toward real life, the sensitivity analysis holds the following implications: Above all, the strong impact of the *initial radius* and the *dispersion factor* is plausible as it suggests that one can actively influence the propagation of respiratory aerosols. Measures that dilute indoor aerosol concentrations, such as ventilation, appear effective. This is in line with scientific investigations on transmission mitigation strategies for SARS-CoV-2 [305] and common practices or recommendations promoted during the COVID-19 pandemic. If ventilation is combined with adequate air exchange, one can achieve a much shorter *half-life* than assumed in my numerical experiments. The experiment with the *half-life* following a log-uniform distribution gives a first indication that the output would react more sensitively to shorter *half-lives*.

The simulations of queues with physical distancing produce sensitivity indices relatively similar to the queues without physical distancing. This does not mean that physical distancing is not effective. It indicates that the model behaves similarly, which is reasonable as long as the model only includes long-range transmission via airborne particles.

When transferring the sensitivity indices to the real world, the reader should bear the model assumptions in mind. For example, the ceiling and the floor confine the aerosol spread in reality, whereas the aerosol clouds in the model expand unhindered in three dimensions. Consequently, the sensitivity index could overestimate the impact of the *dispersion factor* on actual transmission dynamics. This applies similarly to other model simplifications.

To conclude, the sensitivity analysis returns plausible results and, thus, supports the validity of my model. Furthermore, it fosters a better understanding of the model. Since the analysis identifies the *initial radius* and the *dispersion factor* as influential, enhancing the knowledge base and the model for indoor aerosol propagation could effectively reduce the uncertainty in individual exposure levels. Better understanding would help to plan effective measures to mitigate infection risks.

8.4 Monte Carlo forward propagation to quantify exposure

In this section, I present the forward propagation that I carried out in [213] to quantify the variability in the output data for the given input uncertainties. The forward propagation complements the findings of the sensitivity analysis and allows me to draw reliable

conclusions. In the following, I define the simulation settings, explain the outcomes, and discuss the implications.

8.4.1 Simulation set-up for converged summary statistics

Similarly, as for the sensitivity analysis, I verify that the results of the forward propagation approximate the true value well. As explained in Section 8.3.1, this requires a sufficiently large sample size, N. In addition, I need to choose an appropriate number of repetitions, n, of Monte Carlo experiments with different seeds.

I determine the required sample size, N, by running a Monte Carlo simulation exemplary for the single-file queue. Bicher et al. [306] criticize that many simulation studies inadequately treat the convergence of Monte Carlo estimates. The authors emphasize the importance of applying stopping rules for simulations. Being aware of the advantage of such criteria, I adopt a visual approach because basic stopping rules can over- or underestimate the necessary sample size in case of skewed outputs. I run a Monte Carlo simulation with an initial guess of $N = 10^3$ samples. With the given computational resources, the computing time is about 10 min and, hence, acceptable. Subsequently, I compare the empirical distributions resulting from subsets with $N = 10^2$ and $N = 10^3$ samples (see Fig. D.5). The deviation between the empirical distributions is small. Therefore, I deem the output of $N = 10^3$ samples representative.

In a similar manner, I determine the number of required repetitions, n. Based on the experience from the sensitivity analysis (see Section 8.3.1), I define $n = 10$ for the com-

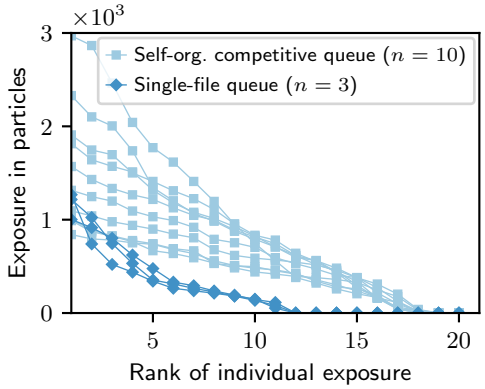


Figure 8.8: The figure shows median exposure levels obtained for repeated Monte Carlo simulations. The quantity of interest is the rank of exposure in descending order. The Monte Carlo experiment for the competitive queue is repeated $n = 10$ times, and that of the single-file queue $n = 3$ times. Note: own figure adapted from [213, p. 12]

petitive queue, $n = 3$ for the single-file queue, and $n = 1$ for the seating scenarios. Fig. 8.8 visualizes variable median values obtained for the repeated and separately evaluated (analogous to option D in Fig. 8.2) Monte Carlo simulations. Here, the median is a better measure of central tendency than the mean because the data is skewed. The different seeds cause variation in the median values of the Monte Carlo simulations. However, the

average of $n = 10$ repetitions for the competitive queue (processed analogously to option C in Fig. 8.2) and $n = 3$ repetitions for the single-file queue (processed analogously to option A in Fig. 8.2) should return representative results.

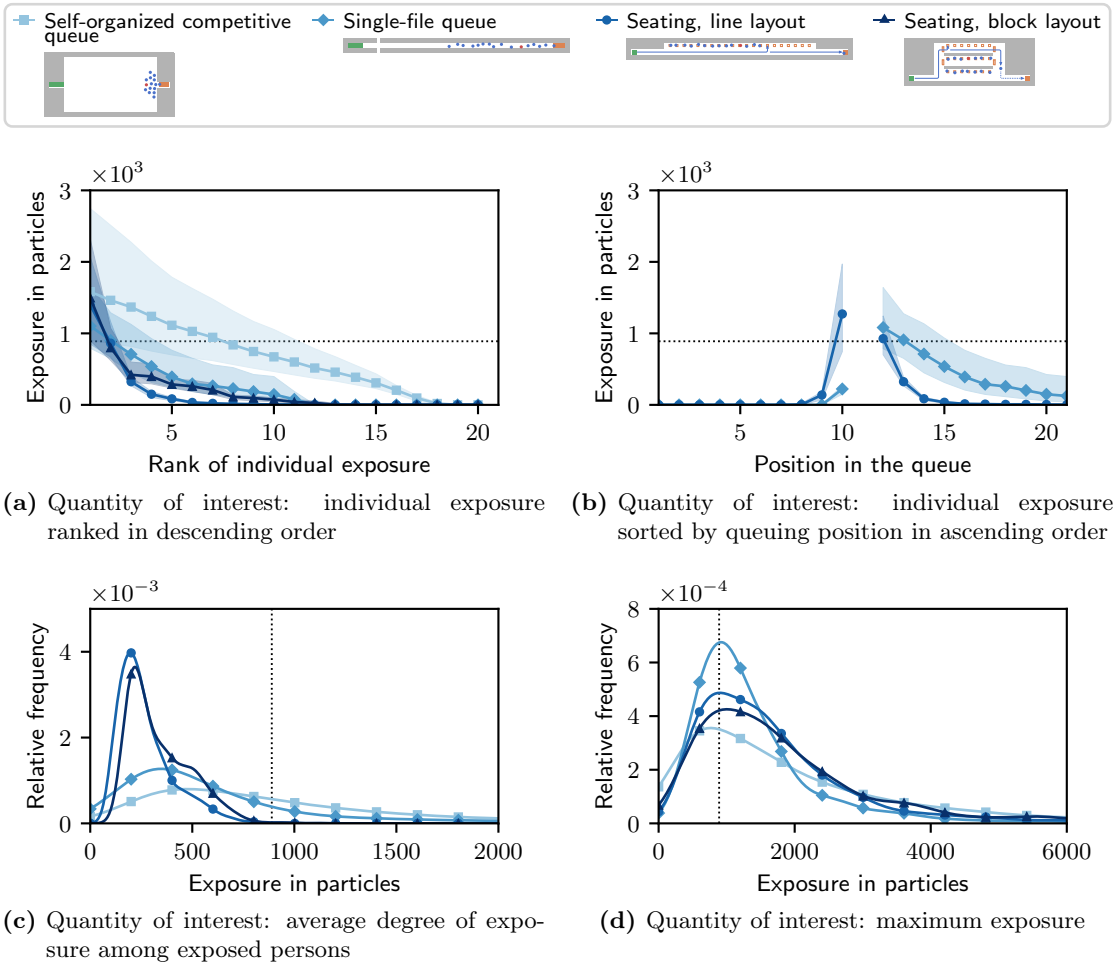
8.4.2 Statistical properties of the output uncertainties

With the simulation set-up defined above, I conduct the Monte Carlo simulations and describe the output statistically. Fig. 8.9 summarizes the results for the scenarios without physical distancing. Most importantly, the comparison to the median exposure obtained for the close contact scenario shows that all scenarios lead to high-risk exposure. Furthermore, the competitive queue causes the most and the highest individual exposure levels. These mainly occur because the waiting people are standing close together. Overall, more persons become exposed simultaneously and for a longer period.

When examining the average exposure (see Fig. 8.9c), one finds lower values in the seating scenarios than in the single-file queue. However, the maximum exposure (see Fig. 8.9d) yields a different picture. Here, the seating scenarios cause slightly higher exposure levels than the single-file queue.

The exposure resolved for each individual (see Fig. 8.9a and Fig. 8.9b) provide more details. The individual exposure levels are characterized by the median to account for the skewness of the data. In particular, Fig. 8.9b demonstrates which persons are most affected in the single-file queue or the seating scenario with the line layout. The general trend of declining exposure in the persons succeeding the infections person is reasonable. The more a person stands toward the end of the queue, the lower the pathogen concentration. Surprisingly, in the seating scenario with the line layout, the person in front of the infectious one experiences higher exposure than the person behind. This occurs in my model because virtual persons find their waiting position not always exactly in the middle of the chair, as is sometimes also the case in reality. The distance between person number 10 and the infectious person is shorter than between person number 12 and the infectious person. Consequently, person number 10 is more often located within the *initial radius* of an aerosol cloud than person number 12. On the one hand, this detail confirms the sensitivity of the exposure at this position with respect to the parameter *half-life*. On the other hand, it demonstrates that one should not over-interpret the median values.

In addition to the median, Fig. 8.9a and Fig. 8.9b also depict the interquartile ranges. The interquartile range is a robust measure of spread in the data while resisting outliers. Here, it marks the highest uncertainty in the output of the competitive queue. The single-file queue varies slightly less, whereas the seating scenarios are generally the least uncertain. The considerable uncertainty in the exposure of the competitive queue mainly stems from uncontrolled conditions where the infectious person is located within the crowd. The uncertainty shrinks considerably if the queue is guided by barriers or fixed waiting positions.



(a) Quantity of interest: individual exposure ranked in descending order

(b) Quantity of interest: individual exposure sorted by queuing position in ascending order

(c) Quantity of interest: average degree of exposure among exposed persons

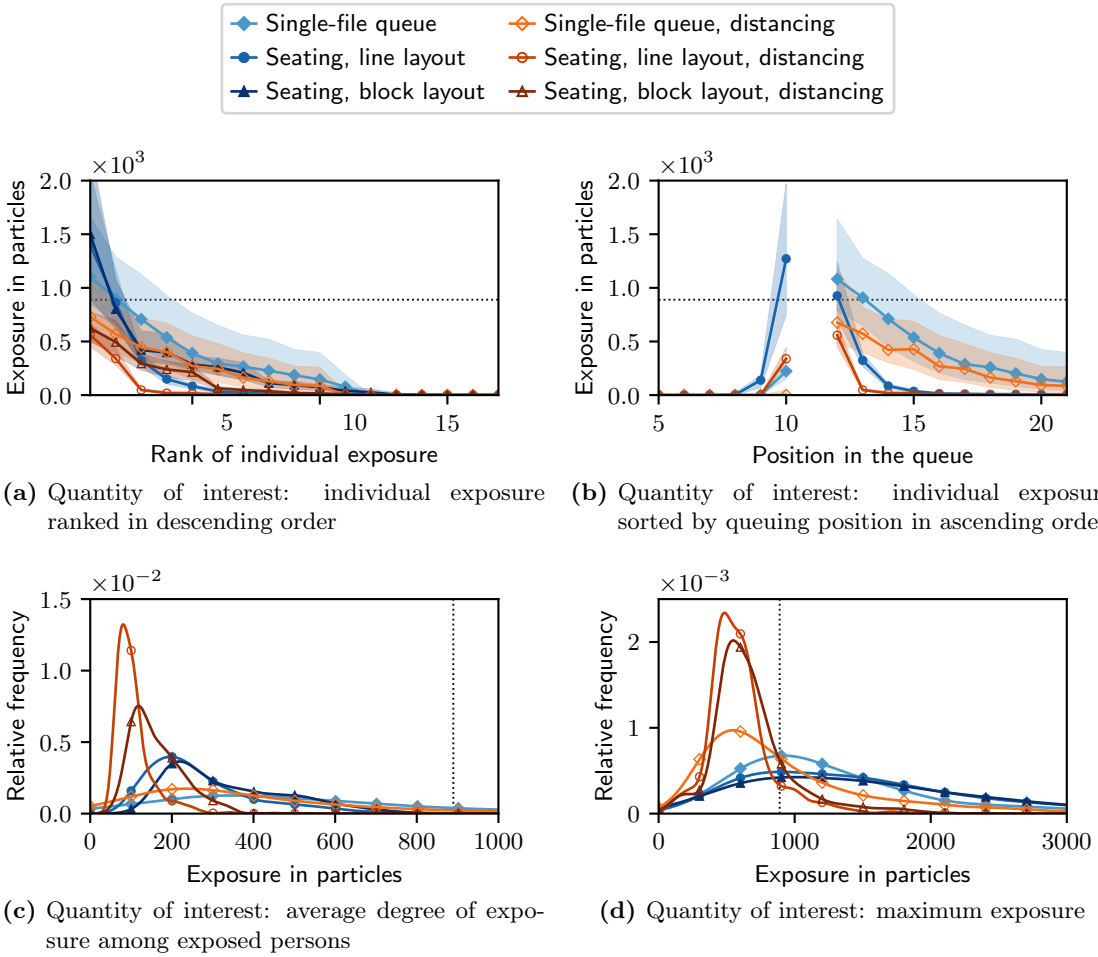
(d) Quantity of interest: maximum exposure

Figure 8.9: The figures summarize the output uncertainty of four quantities of interest as a result of evaluating $N = 10^3$ (pseudo-)random Monte Carlo samples. The individual exposure levels in (a) and (b) are described by the median (line) and the interquartile range (shaded area). The exposure for the infectious person at rank 21 in (a) or position 10 in (b) is undefined. The distributions in (c) and (d) are represented by Kernel Density Estimators. The dotted lines correspond to the median exposure of the close contact scenario. Note: own figures adapted from [213, p. 10].

8.4.3 Impact of physical distancing on the exposure risk

I now also take into account physical distancing. The results are shown in Fig. 8.10. Compared to the scenarios without physical distancing, the median values in Fig. 8.10a and Fig. 8.10b show the same trend. However, the median exposure drops to a lower level below the median reference exposure of a close contact. The uncertainty in the individuals' exposure shrinks slightly. The distributions of the average exposure (Fig. 8.10c) and the maximum exposure (Fig. 8.10d) peak at lower values and indicate smaller

variations in the data. Overall, the seating scenario with a line layout results in the fewest and lowest exposure levels.



(a) Quantity of interest: individual exposure ranked in descending order

(b) Quantity of interest: individual exposure sorted by queuing position in ascending order

(c) Quantity of interest: average degree of exposure among exposed persons

(d) Quantity of interest: maximum exposure

Figure 8.10: The figures compare the output uncertainty of queuing scenarios without physical distancing (shades of blue) and with physical distancing (shades of orange). The individual exposure levels in (a) and (b) are described by the median (line) and the interquartile range (shaded area). The distributions in (c) and (d) are represented by Kernel Density Estimators. The dotted lines correspond to the median exposure of the close contact scenario. Note: own figure adapted from [213, p. 14].

8.4.4 Implications of the uncertainty analysis

The above findings of the Monte Carlo experiments have implications. Keeping in mind that my simulations focus on long-range transmission via airborne particles, they shed light on the following aspects of SARS-CoV-2 transmission:

First of all, the forward propagation results support the hypothesis stated in Chapter 7 that the queuing behavior influences the exposure risk. They demonstrate that the

self-organized queue represents the most critical case concerning the average exposure, the maximum exposure, and the number of exposed persons. The other queues guided by explicit queue management clearly show fewer and lower exposure levels. Among these actively managed queues, the seating scenarios, on average, show slightly better results than the single-file queue. While such fine differences may be scenario-specific, I generally conclude that crowd management can reduce the exposure risk. Physical distancing may contribute to this. However, the reduction seems to be linked tighter to the queuing pattern. Hence, the simulations are generally in line with the transmission mitigation strategy of explicitly organizing queues during the COVID-19 pandemic.

The Monte Carlo experiments also highlight the importance of taking uncertainties into account. Reliable simulation results and their interpretation inevitably require information about the output variability. The interquartile ranges, used as a measure of statistical dispersion, are narrow in most cases. Thus, the ranking of the scenarios becomes much more trustworthy than the initial guess based on point estimates.

Moreover, the uncertainty analysis complements the Sobol' sensitivity analysis. It offers a quantitative description of the exposure and, thus, enables us to better understand the sensitivity. Importantly, the uncertainty analysis puts the sensitivities obtained separately for each scenario in context. The outputs obtained from the competitive queue and the single-file queue vary more than those of the seating scenarios. This means that the competitive queue and the single-file queue react more sensitively to the same input variations than the other two scenarios. Hence, the queuing pattern is a third influential parameter besides the *initial radius* and the *dispersion factor*. This would become apparent immediately if the queuing strategy were included in the uncertain parameter space. However, the differences between the scenarios are challenging to characterize in terms of a parameter.

8.5 Summary

I conducted a global sensitivity analysis with the Sobol' method and forward propagation experiments based on Monte Carlo simulations. The aim was to quantify the effect of uncertainties in model parameters on the exposure of individuals in different scenarios. I analyzed the parameters related to the decay and spatiotemporal propagation of aerosol clouds (*half-life*, *initial radius*, and *dispersion factor*). The scenarios, first introduced in Chapter 7, represent situations where one infectious person and several susceptible persons form a self-organized queue or follow an actively managed queue to pass a service unit.

The sensitivity analysis showed that the parameters determining the spatiotemporal propagation of the aerosol clouds dominate the variability in the output. Consequently, improving the knowledge base to reduce the input uncertainty or refining the model with respect to aerosol propagation would effectively improve the reliability of the simulation. Regarding practical application, the high sensitivity indicates that measures accelerating the dilution of aerosols are effective. The remaining uncertain parameter related to the

decay of the airborne pathogen load has no significant influence and, therefore, could be fixed in future simulations.

The Monte Carlo simulations revealed that queue management positively impacts the exposure risk. Average exposure levels and variability in the individuals' exposure decreased substantially in the actively managed queues. Physical distancing added to this, but the effect of queue management was larger. Seeing that the model merely focuses on transmission via smaller aerosol particles, this appears reasonable and supports the application of tailored actions against airborne pathogen transmission.

Overall, the sensitivity analysis and the forward propagation returned plausible trends and, thus, support the model's validity. They showed exemplary how to generate trustworthy outputs in the presence of large parameter uncertainties. Importantly, quantifying uncertainties is possible because my model is parsimonious. Uncertainty quantification methods, even more efficient ones, quickly become infeasible for computationally demanding models. This emphasizes the advantage of more straightforward modeling approaches given the various unknowns related to airborne transmission of pathogens, particularly SARS-CoV-2.

9 Summary, conclusion, and future directions

This chapter summarizes the key research findings, evaluates my achievements, and addresses their limitations. Finally, I suggest directions for further research.

9.1 Findings summary

This dissertation investigated the overarching question: *How can airborne transmission of pathogens such as SARS-CoV-2 be modeled and simulated?* The objective was to derive a mathematical model for airborne pathogen transmission between individuals and to translate it into a simulation program that researchers from different disciplines can use for and adapt to their own research. To demonstrate the validity of the model and how one can obtain reliable results despite input uncertainties, I reenacted superspreading events and predicted individual exposure risks of people gathering indoors.

Developing and using such a model requires interdisciplinary knowledge. Therefore, Part I of this work provided the essential background about SARS-CoV-2 transmission, current approaches to modeling inter-individual transmission, and mathematical methods designed to improve the reliability of simulations suffering from uncertainties. I sought to answer the following sub-questions:

What are the essential characteristics of SARS-CoV-2 transmission? In Chapter 2, I contextualized inconsistent findings in the literature about the spread of the virus and extracted that it is mainly transmitted via inhalation of respiratory aerosols. As a dominant transmission mechanism, breathing produces primarily smaller respiratory aerosol particles. In indoor spaces without adequate ventilation, these remain airborne long enough to cause infection in individuals that are not in close proximity to the infectious source. However, despite extensive efforts of the scientific community, many parameters remain uncertain. Above all, a biologically plausible dose-response relationship describing the infection probability in relation to the dose taken in has not yet been established.

How does the literature from infectious disease modeling operationalize these mechanisms, and which facets are not adequately represented? Chapter 3 reviewed modeling approaches adopted at global to local scales. At the local scale, between-host transmission models typically assume either a homogeneously mixed pathogen concentration within a room or inhomogeneous conditions. State-of-the-art models define transmission as a function of the proximity to the infectious person. However, this fails to capture long-range transmission of airborne pathogens. Motivated by the outbreak of the COVID-19 pandemic in 2020, the concept of a transmission medium, such as aerosol particles, emerged in recent developments. First attempts pursuing this direction struggled with finding an adequate level of granularity between the model aspects of transmission and crowd dynamics. At the time of writing this dissertation, they are still in a development stage, so none of these has been established. I addressed this scientific gap.

How should uncertainties about airborne pathogen transmission be addressed in modeling and simulation? Chapter 4 introduced mathematical methods for quantifying uncertainties and, thus, for improving the reliability of simulation outcomes. Currently, there are only few simulation studies of individual-based SARS-CoV-2 transmission that treat uncertainties in the model parameters. For my own simulations carried out in the second part of this work, I identified the following two methods as relevant: Global sensitivity analysis based on the Sobol' method to quantify the importance of uncertain model input parameters and Monte Carlo simulations to quantify output uncertainties.

Based on this background knowledge, Part II continued with my own contribution of developing a model and running simulations. This content is partly covered by the initial proposal of my model in [17], a brief description of the software [212], and a parameter study conducted in [213]. My dissertation tied these previous works together and updated the simulations presented in [17]. I pursued the following concrete questions:

How can important aspects of inter-individual transmission of SARS-CoV-2 via airborne respiratory aerosols be described mathematically? As detailed in Chapter 5, the first step was to model transmission in unventilated indoor environments. My formulation integrates respiratory activity into an individual-based model for human motion, a so-called microscopic crowd model. This enabled me to define pathogen transmission among virtual persons moving around in arbitrarily designed rooms or buildings. Infectious persons exhale pathogens bound to aerosol clouds, whereas susceptible persons accumulate pathogens inhaled from these aerosol clouds. The aerosol clouds increase spatially with time in all directions. Airflows caused, for example, by ventilation or humans dissipating thermal energy are not explicitly modeled. Deactivation and sedimentation of viral load is represented by an exponential decay. I deliberately chose such a coarse definition to obtain a compatible level of detail between crowd model and transmission model.

How can the mathematical model be implemented efficiently as sustainable software in the sense of reusable and verified code? Chapter 6 presented my implementation of the model as a simulation program. I integrated the model into Vadere, an established open-source simulation framework for crowd dynamics, using a modular architecture with clearly separated components. I created flexibly adaptable code through parametrization and achieved a suitable level of verification through automated testing. Thus, I ensured that the extension is reusable and reliable. By implementing an additional model for exposure via proximity, I also demonstrated that my model can easily be extended or replaced by alternative approaches.

How can one reliably predict the exposure risk for everyday situations? In Chapter 7, I described how to predict and compare the exposure risk occurring in several indoor situations. To that end, I simulated a situation of a close contact between an infectious person and a susceptible person, which is commonly associated with a high risk of infection. Hence, I obtained a reference value for high-risk exposure. This reference value allowed me to evaluate the relative criticality of other scenarios despite the absence of a dose-response model for infection risk assessment. Reenacting two COVID-19 superspreading events that occurred in a restaurant and during a choir rehearsal showed that

the model produces plausible results. Hence, these simulations helped to build trust in the model. In addition, I compared the exposure risk between several settings where people wait in the same room for a given time and a situation where they walk along an indoor corridor. The waiting people experienced substantially higher exposure than the ones in the corridor scenario. The simulations were influenced by considerable variations in the input parameters, which called for quantifying uncertainties to obtain reliable results.

How can uncertainties in the simulation be quantified? I quantified the output uncertainty in Chapter 8 through sensitivity analysis and forward propagation. I found that uncertainties concerning the spatiotemporal spread of aerosol clouds have a large impact on the individuals' exposure. This means reducing these uncertainties would effectively improve the reliability of the simulation. In contrast, variability in the parameter related to the decay of the pathogen load barely influenced the exposure. Therefore, this parameter can be fixed in the future. Comparing the exposure levels associated with different movement patterns demonstrated that crowd dynamics also affect the exposure. More precisely, the results showed that enforcing a single-file queue or establishing a ticket system reduced the exposure risks compared to unorganized queuing. Thus, I delivered a blueprint how users of my model can estimate exposure risks and identify effective transmission mitigation strategies.

9.2 Conclusion

My work contributes to closing the research gap of modeling airborne SARS-CoV-2 transmission between individuals in indoor spaces. This is important to better understand the exposure risks in specific situations of daily life, for example, public transportation where people pass through a waiting area or a corridor. Simulating and comparing such scenarios complements state-of-the-art risk assessments. In particular, it reveals individual exposure levels instead of an average risk. The most important aspects of my contribution comprise:

- A model with a suitable resolution of the transmission mechanisms, taking into account airborne pathogen spread and human movement;
- An open-source simulation tool that enables me and others to analyze specific use cases;
- Exemplary risk assessments conducted with this tool and insights into the criticality of specific scenarios and their implications.

Achieving this entailed, above all, the challenge of combining the involved disciplines at a suitable level of granularity. Each of the research areas offers sophisticated approaches to solve sub-problems, such as describing aerosol generation and propagation or human movement. Respiratory aerosols do not instantaneously distribute homogeneously within enclosed spaces after being exhaled. Therefore, a person's exposure to airborne pathogens depends on the position of both the pathogens as well as the person.

Consequently, it is important to consider that we humans interact with our environment and move around. I used a tested and validated microscopic crowd model to capture human locomotion in 2D. Since my exposure model should match this level of detail, I represented exhalation of aerosols, persistence of airborne pathogens, and their respiratory deposition in a simple manner. The model can and should be refined step-wise when more knowledge becomes available in the future. Compared to high-fidelity airflow models that include moving persons, my representation of aerosol propagation is intentionally coarse and, consequently, computationally parsimonious. This allowed me to run numerous simulations and, thus, to obtain outputs not just for one specific but various conditions. Moreover, a less complex model is comprehensible for a broader audience and supports interdisciplinary research.

To foster such collective efforts, I implemented my model according to state-of-the-art criteria for research software. The simulation tool represents another important accomplishment in my work. The corresponding implementation is integrated into an open-source crowd simulator. I ensured and demonstrated in an exemplary way that my new feature offers adequate and easy-to-use interfaces for alternative modeling approaches. Thus, others can use or extend the tool as needed. Importantly, the manageable amount of six parameters related to the exposure model offers uncomplicated and already wide-ranging adaptations, for example, to similar diseases. The underlying crowd model adds additional flexibility, which I used to simulate various indoor scenarios.

I conducted exemplary numerical experiments with this tool. Since the empirical data that is currently available does not suffice to estimate the dose-response, calculating infection risks was not possible. Instead, I compared the exposure occurring in the scenarios to obtain a relative ranking of the criticality. As with any simulation of SARS-CoV-2 transmission, there are several unknowns. Therefore, I applied uncertainty quantification methods and improved the reliability of my simulations in the presence of variations in the input. Here, the parsimony of my model came in handy because such methods quickly become infeasible for computationally demanding models. My simulation studies showed that quantifying uncertainties is inevitable to obtain trustworthy and meaningful results.

It should be noted that these simulation results have limitations. The uncertainty quantification indicated that the spatiotemporal spread of aerosols has a strong impact on the individuals' exposure. Acquiring further knowledge to reduce the input uncertainty and refining the model in this regard could effectively improve the reliability. However, one should keep the resolution of the crowd model and the transmission model balanced when introducing more detailed effects, such as ventilation. Furthermore, this work focused on SARS-CoV-2 transmission via inhalation of airborne particles. If the aim is to transfer the model to other diseases, it could be necessary to include transmission via fomites or respiratory particles that follow ballistic trajectories.

Despite these limitations, my findings can be applied to real-world scenarios. The simulation framework I extended is capable of reenacting various everyday situations. In principle, this enables anyone to assess the exposure risk for specific use cases. Indeed, I am in contact with external users who are testing the model in the context of airplane

boarding. My own exemplary simulation studies illustrated in an accessible way that critical situations are likely to occur if an infectious person stays in place for a certain time. In contrast, passing or walking normally behind a COVID-19 case results in low exposure to airborne pathogens. Such information could relief fears and support practitioners in taking informed decisions. The simulations showcased that appropriate crowd management can effectively reduce the exposure risk. Overall, I successfully modeled and simulated between-host transmission of SARS-CoV-2 in dynamic crowds.

9.3 Future directions

One of the main limitations of my model is that, while the virtual persons can breathe, they cannot yet speak, sneeze, or cough. To capture such expiratory events, one needs to account for transmission via larger aerosol particles that follow ballistic trajectories. When considering particle diameters below 5 µm, speaking produces a similar particle size distribution as breathing (see Fig. 2.4). Therefore, the smaller particles expelled through speaking can be captured by aerosol clouds as described in Chapter 5 (see Fig. 5.2). However, the particle size distribution for speaking peaks also at diameters of around 50 µm. To capture this second mode, one can introduce a transmission medium, complementing the aerosol clouds. Encouraged by the approach in [191], I propose to use a circular sector oriented in the heading direction of a person to represent the emission of larger aerosol particles (see Fig. 9.1). I implemented a first draft of this concept in the Vadere class `Droplets`, so others can test the prototype.[1] The pathogen load, frequency of occurrence, persistence, central angle, and radius of the segment remain to be determined. Depending on the chosen parameter values, this reflects different respiratory events. In a further step, one could account for shielding effects, exhalation and inhalation at different heights, and the head orientation of the inhaling person. Such a concept would match the level of granularity of the current transmission model.

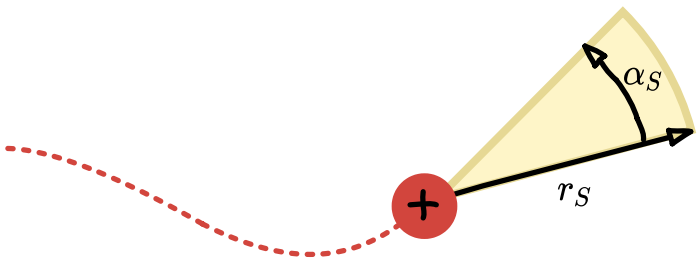


Figure 9.1: Draft of a transmission model accounting for larger aerosol particles: An infectious person (red) moves from left to right (dotted line) and emits larger respiratory aerosol particles that follow a ballistic trajectory. The larger particles reach the area covered by the circular sector (yellow) with central angle α_S and radius r_S. Other persons standing inside that area would become exposed.

[1] `https://gitlab.lrz.de/vadere/vadere`, commit 8d6fb02a, accessed on April 2, 2024

The sensitivity analysis in Chapter 8 indicated that parameter uncertainties related to the spread of aerosol clouds have a high impact on the output uncertainty. Consequently, it seems advisable to refine the representation of aerosol clouds and to collect more data to reduce the uncertainty of the associated input parameters. For example, one could learn aerosol propagation from image data. This also might enable us to improve the underlying model for aerosol generation and spread.

Finally, since validation is a difficult task in modeling SARS-CoV-2 transmission, I propose to investigate the number of inhaled particles experimentally to obtain a ground truth. According to aerosol researchers I and the Vadere development team consulted, one could artificially generate labeled aerosols and let a person, equipped with a inhalable aerosol sampler, pass the aerosol. The sampler would reveal the number of particles that deposit in the lungs. This would help us to compare simulated to actual exposure levels and, hence, to improve the validation.

There are promising opportunities for further developments. I hope that my work fosters interdisciplinary research, contributes to aerosols being acknowledged as essential for transmission of certain pathogens, and adds to the maturation of small-scale approaches in infectious disease modeling.

References

[1] C. J. Carlson, G. F. Albery, C. Merow, C. H. Trisos, C. M. Zipfel, E. A. Eskew, K. J. Olival, N. Ross, and S. Bansal. Climate change increases cross-species viral transmission risk. *Nature*, 607(7919):555–562, 2022. doi:10.1038/s41586-022-04788-w.

[2] T. Wu, C. Perrings, A. Kinzig, J. P. Collins, B. A. Minteer, and P. Daszak. Economic growth, urbanization, globalization, and the risks of emerging infectious diseases in China: A review. *Ambio*, 46(1):18–29, 2017. doi:10.1007/s13280-016-0809-2.

[3] P. K. Betz, J. Stoll, V. Grappendorf, J. Gilg, M. Zeumer, M. Klitz, L. Spataro, A. Klein, L. Rothenhaeusler, H. Bohnacker, H. Kraemer, M. Meyer-Hermann, S. Somogyi, A. Gerndt, and M. J. Kühn. ESID: Exploring the Design and Development of a Visual Analytics Tool for Epidemiological Emergencies. In *2023 IEEE VIS Workshop on Visualization for Pandemic and Emergency Responses (Vis4PandEmRes)*. IEEE, 2023. doi:10.1109/vis4pandemres60343.2023.00007.

[4] M. T. Madigan, J. M. Martinko, K. S. Bender, D. H. Buckley, and D. A. Stahl. *Brock Biology of Microorganisms*. Pearson Education, Limited, Boston, 14th edition, 2014. ISBN 9781292018317.

[5] B. M. Althouse, E. A. Wenger, J. C. Miller, S. V. Scarpino, A. Allard, L. Hébert-Dufresne, and H. Hu. Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control. *PLOS Biology*, 18(11): e3000897, 2020. doi:10.1371/journal.pbio.3000897.

[6] W. F. Wells. *Airborne contagion and air hygiene*. Cambridge, Mass., Harvard Univ. Press, 1955.

[7] E. C. Riley, G. Murphy, and R. L. Riley. Airborne spread of measles in a suburban elementary school. *American Journal of Epidemiology*, 107(5):421–432, 1978. doi:10.1093/oxfordjournals.aje.a112560.

[8] C. Noakes and A. Sleigh. Applying the Wells-Riley equation to the risk of airborne infection in hospital environments: The importance of stochastic and proximity effects. In *Indoor Air 2008 : The 11th International Conference on Indoor Air Quality and Cl Indoor Air 2008, 17–22nd August 2008, Copenhagen, Denmark.* The University of Leeds, 2008. URL `https://eprints.whiterose.ac.uk/7702/`. Accessed on April 5, 2024.

[9] G. N. Sze To and C. Y. H. Chao. Review and comparison between the Wells-Riley and dose-response approaches to risk assessment of infectious respiratory diseases. *Indoor Air*, 20(1):2–16, 2010. doi:10.1111/j.1600-0668.2009.00621.x.

[10] A. Aganovic and E. Kadric. Does the exponential Wells-Riley model provide a good fit for human coronavirus and rhinovirus? A comparison of four dose-response models based on human challenge data. *Risk Analysis*, 2023. doi:10.1111/risa.14178.

[11] A. Johansson and L. Goscé. Utilizing Crowd Insights to Refine Disease-Spreading Models. In U. Weidmann, U. Kirsch, and M. Schreckenberg, editors, *Pedestrian and Evacuation Dynamics 2012*, pages 1395–1403. Springer, Cham, 2014. doi:10.1007/978-3-319-02447-9_116.

[12] V. Vuorinen, M. Aarnio, M. Alava, V. Alopaeus, N. Atanasova, M. Auvinen, N. Balasubramanian, H. Bordbar, P. Eröstö, R. Grande, N. Hayward, A. Hellsten, S. Hostikka, J. Hokkanen, O. Kaario, A. Karvinen, I. Kivistö, M. Korhonen, R. Kosonen, J. Kuusela, S. Lestinen, E. Laurila, H. J. Nieminen, P. Peltonen, J. Pokki, A. Puisto, P. Råback, H. Salmenjoki, T. Sironen, and M. Österberg. Modelling aerosol transport and virus exposure with numerical simulations in relation to SARS-CoV-2 transmission by inhalation indoors. *Safety Science*, 130: 104866, 2020. doi:10.1016/j.ssci.2020.104866.

[13] T. Antczak, B. Skorupa, M. Szurlej, R. Weron, and J. Zabawa. Simulation Modeling of Epidemic Risk in Supermarkets: Investigating the Impact of Social Distancing and Checkout Zone Design. In M. Paszynski, D. Kranzlmüller, V. V. Krzhizhanovskaya, J. J. Dongarra, and P. M. A. Sloot, editors, *Computational Science - ICCS 2021*, volume 12742 of *Lecture Notes in Computer Science*, pages 26–33. Springer International Publishing, 2021. doi:10.1007/978-3-030-77961-0_3.

[14] J. Wang, H. Tang, J. Wang, and Z. Zhong. An agent-based study on the airborne transmission risk of infectious disease in a fever clinic during COVID-19 pandemic. *Building and Environment*, 218:109118, 2022. doi:10.1016/j.buildenv.2022.109118.

[15] H. Salmenjoki, M. Korhonen, A. Puisto, V. Vuorinen, and M. J. Alava. Modelling aerosol-based exposure to SARS-CoV-2 by an agent based Monte Carlo method: Risk estimates in a shop and bar. *PLOS ONE*, 16(11):e0260237, 2021. doi:10.1371/journal.pone.0260237.

[16] S. Rahn, M. Gödel, and G. Köster. Towards a model for SARS-CoV-2 transmission at a local scale. In *Proceedings of the 10th International Conference on Pedestrian and Evacuation Dynamics (PED2021): Melbourne, Australia, November 29–30, 2021*, page 20, 2021. URL `https://drive.google.com/file/d/1OdQIxcxaoCuQJUjQE4AC0NJ_qzC4mVHH/view`. Accessed on April 5, 2024.

[17] S. Rahn, M. Gödel, G. Köster, and G. Hofinger. Modelling airborne transmission of SARS-CoV-2 at a local scale. *PLOS ONE*, 17(8):1–24, 2022. doi:10.1371/journal.pone.0273820.

[18] A. Grignard, T. Nguyen-Huu, P. Taillandier, L. Alonso, N. Ayoub, M. Elkatsha, G. Palomo, M. Gomez, M. Siller, M. Gamboa, C. I. Moreno, and K. Larson. Using Agent-Based Modelling to Understand Advantageous Behaviours Against COVID-19 Transmission in the Built Environment. In K. H. Van Dam and N. Verstaevel, editors, *Multi-Agent-Based Simulation XXII*, Lecture Notes in Computer Science, pages 86–98. Springer, Cham, 2022. doi:10.1007/978-3-030-94548-0_7.

[19] D. Mukherjee and G. Wadhwa. A mesoscale agent based modeling framework for flow-mediated infection transmission in indoor occupied spaces. *Computer Methods in Applied Mechanics and Engineering*, 401:115485, 2022. doi:10.1016/j.cma.2022.115485.

[20] Robert Koch-Institut. Kontaktpersonen-Nachverfolgung (KP-N) bei SARS-CoV-2-Infektionen, Stand 14.1.2022, außer Kraft seit 2.5.2022. Online, 2022. URL `https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Kontaktperson/Management.html`. Accessed on April 5, 2024.

[21] M. Haghani and M. Sarvi. Crowd model calibration at strategic, tactical, and operational levels: Full-spectrum sensitivity analyses show bottleneck parameters are most critical, followed by exit-choice-changing parameters. *Transportation Letters*, 0(0):1–28, 2023. doi:10.1080/19427867.2023.2195729.

[22] N. H. Fefferman, J. S. McAlister, B. S. Akpa, K. Akwataghibe, F. T. Azad, K. Barkley, A. Bleichrodt, M. J. Blum, L. Bourouiba, Y. Bromberg, K. S. Candan, G. Chowell, E. Clancey, F. A. Cothran, S. N. DeWitte, P. Fernandez, D. Finnoff, D. T. Flaherty, N. L. Gibson, N. Harris, Q. He, E. T. Lofgren, D. L. Miller, J. Moody, K. Muccio, C. L. Nunn, M. Papes, I. C. Paschalidis, D. K. Pasquale, J. M. Reed, M. B. Rogers, C. L. Schreiner, E. B. Strand, C. S. Swanson, H. L. Szabo-Rogers, and S. J. Ryan. A New Paradigm for Pandemic Preparedness. *Current Epidemiology Reports*, 10(4):240–251, 2023. doi:10.1007/s40471-023-00336-w.

[23] H.-J. Bungartz, S. Zimmer, M. Buchholz, and D. Pflüger. *Modeling and Simulation: An Application-Oriented Introduction*. Springer Undergraduate Texts in Mathematics and Technology. Springer, Berlin Heidelberg, 2014. doi:10.1007/978-3-642-39524-6.

[24] S. Schlesinger. Terminology for model credibility. *Simulation*, 32(3):103–104, 1979. doi:10.1177/003754977903200304.

[25] Norddeutscher Rundfunk. Das Coronavirus-Update von NDR Info. Online, 2023. URL `https://www.ndr.de/nachrichten/info/podcast4684.html`. Accessed on April 5, 2024.

[26] Robert Koch-Institut. Epidemiologischer Steckbrief zu SARS-CoV-2 und COVID-19, Stand 26.11.2021. Online, 2021. URL `https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html`. Accessed on April 5, 2024.

[27] World Health Organization. Listings of WHO's response to COVID-19. Online, 2021. URL https://www.who.int/news/item/29-06-2020-covidtimeline. Accessed on April 5, 2024.

[28] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, 5(4):536–544, 2020. doi:10.1038/s41564-020-0695-z.

[29] World Health Organization. Novel Coronavirus (2019-nCoV) Situation report – 22. Online, 2020. URL https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf. Accessed on April 5, 2024.

[30] World Health Organization. WHO Director-General's opening remarks at the mediabriefing on COVID-19 - 11 March 2020. Online, 2020. URL https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. Accessed on April 5, 2024.

[31] World Health Organization. WHO COVID-19 Dashboard. Online, 2020. URL https://covid19.who.int/. Accessed on August 10, 2023.

[32] World Health Organization. Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic. Online, 2023. URL https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic. Accessed on April 5, 2024.

[33] International Committee on Taxonomy of Viruses. The International Code of Virus Classification and Nomenclature (ICVCN). Online, 2021. URL https://ictv.global/about/code. Accessed on April 5, 2024.

[34] S. Payne. Family Coronaviridae. In *Viruses*, pages 149–158. Elsevier, 2017. doi:10.1016/b978-0-12-803109-4.00017-9.

[35] J. D. Almeida, D. M. Berry, C. H. Cunningham, D. Hamre, M. S. Hofstad, L. Mallucci, K. McIntosh, and D. A. J. Tyrrell. Virology: Coronaviruses. *Nature*, 220 (5168):650–650, 1968. doi:10.1038/220650b0.

[36] P. C. Y. Woo, S. K. P. Lau, C. S. F. Lam, C. C. Y. Lau, A. K. L. Tsang, J. H. N. Lau, R. Bai, J. L. L. Teng, C. C. C. Tsang, M. Wang, B.-J. Zheng, K.-H. Chan, and K.-Y. Yuen. Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene

Source of Gammacoronavirus and Deltacoronavirus. *Journal of Virology*, 86(7): 3995–4008, 2012. doi:10.1128/jvi.06540-11.

[37] H. M. Ashour, W. F. Elkhatib, M. M. Rahman, and H. A. Elshabrawy. Insights into the Recent 2019 Novel Coronavirus (SARS-CoV-2) in Light of Past Human Coronavirus Outbreaks. *Pathogens*, 9(3):186, 2020. doi:10.3390/pathogens9030186.

[38] Z. Zhu, X. Lian, X. Su, W. Wu, G. A. Marraro, and Y. Zeng. From SARS and MERS to COVID-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respiratory Research*, 21(1), 2020. doi:10.1186/s12931-020-01479-w.

[39] J. Gu. Human Coronavirus Structure. Online, 2020. URL `https://app.biorender.com/biorender-templates/figures/5e99f5395fd61e0028682c01/t-5f21e90283765600b08fbe9d-human-coronavirus-structure`. Accessed on April 5, 2024.

[40] S. S. Manathunga, I. A. Abeyagunawardena, and S. D. Dharmaratne. A comparison of transmissibility of SARS-CoV-2 variants of concern. *Virology Journal*, 20(1), 2023. doi:10.1186/s12985-023-02018-x.

[41] S. J. R. da Silva, J. C. F. do Nascimento, R. P. G. Mendes, K. M. Guarines, C. T. A. da Silva, P. G. da Silva, J. J. F. de Magalhães, J. R. J. Vigar, A. Silva-Júnior, A. Kohl, K. Pardee, and L. Pena. Two Years into the COVID-19 Pandemic: Lessons Learned. *ACS Infectious Diseases*, 8(9):1758–1814, 2022. doi:10.1021/acsinfecdis.2c00204.

[42] J. O. Wrenn, S. B. Pakala, G. Vestal, M. H. Shilts, H. M. Brown, S. M. Bowen, B. A. Strickland, T. Williams, S. A. Mallal, I. D. Jones, J. E. Schmitz, W. H. Self, and S. R. Das. COVID-19 severity from Omicron and Delta SARS-CoV-2 variants. *Influenza and Other Respiratory Viruses*, 16(5):832–836, 2022. doi:10.1111/irv.12982.

[43] R. da Rosa Mesquita, L. C. F. S. Junior, F. M. S. Santana, T. F. de Oliveira, R. C. Alcântara, G. M. Arnozo, E. R. da Silva Filho, A. G. G. dos Santos, E. J. O. da Cunha, S. H. S. de Aquino, and C. D. F. de Souza. Clinical manifestations of COVID-19 in the general population: systematic review. *Wiener klinische Wochenschrift*, 133(7-8):377–382, 2021. doi:10.1007/s00508-020-01760-4.

[44] X. Qiu, A. I. Nergiz, A. E. Maraolo, I. I. Bogoch, N. Low, and M. Cevik. The role of asymptomatic and pre-symptomatic infection in SARS-CoV-2 transmission – a living systematic review. *Clinical Microbiology and Infection*, 27(4):511–519, 2021. doi:10.1016/j.cmi.2021.01.011.

[45] J. M. Bartleson, D. Radenkovic, A. J. Covarrubias, D. Furman, D. A. Winer, and E. Verdin. SARS-CoV-2, COVID-19 and the aging immune system. *Nature Aging*, 1(9):769–782, 2021. doi:10.1038/s43587-021-00114-7.

[46] B. Killingley, A. J. Mann, M. Kalinova, A. Boyers, N. Goonawardane, J. Zhou, K. Lindsell, S. S. Hare, J. Brown, R. Frise, E. Smith, C. Hopkins, N. Noulin, B. Löndt, T. Wilkinson, S. Harden, H. McShane, M. Baillet, A. Gilbert, M. Jacobs, C. Charman, P. Mande, J. S. Nguyen-Van-Tam, M. G. Semple, R. C. Read, N. M. Ferguson, P. J. Openshaw, G. Rapeport, W. S. Barclay, A. P. Catchpole, and C. Chiu. Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. *Nature Medicine*, 28(5):1031–1041, 2022. doi:10.1038/s41591-022-01780-9.

[47] L. M. Brosseau, K. Escandón, A. K. Ulrich, A. L. Rasmussen, C. J. Roy, G. J. Bix, S. V. Popescu, K. A. Moore, and M. T. Osterholm. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Dose, Infection, and Disease Outcomes for Coronavirus Disease 2019 (COVID-19): A Review. *Clinical Infectious Diseases*, 75(1):e1195–e1201, 2022. doi:10.1093/cid/ciab903.

[48] S. Karimzadeh, R. Bhopal, and H. N. Tien. Review of infective dose, routes of transmission and outcome of COVID-19 caused by the SARS-COV-2: comparison with other respiratory viruses– CORRIGENDUM. *Epidemiology and Infection*, 149:e116, 2021. doi:10.1017/s0950268821001084.

[49] A. Popa, J.-W. Genger, M. D. Nicholson, T. Penz, D. Schmid, S. W. Aberle, B. Agerer, A. Lercher, L. Endler, H. Colaço, M. Smyth, M. Schuster, M. L. Grau, F. Martínez-Jiménez, O. Pich, W. Borena, E. Pawelka, Z. Keszei, M. Senekowitsch, J. Laine, J. H. Aberle, M. Redlberger-Fritz, M. Karolyi, A. Zoufaly, S. Maritschnik, M. Borkovec, P. Hufnagl, M. Nairz, G. Weiss, M. T. Wolfinger, D. von Laer, G. Superti-Furga, N. Lopez-Bigas, E. Puchhammer-Stöckl, F. Allerberger, F. Michor, C. Bock, and A. Bergthaler. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Science Translational Medicine*, 12(573):eabe2555, 2020. doi:10.1126/scitranslmed.abe2555.

[50] M. A. Martin and K. Koelle. Comment on "Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2". *Science Translational Medicine*, 13(617), 2021. doi:10.1126/scitranslmed.abh1803.

[51] R. Sender, Y. M. Bar-On, S. Gleizer, B. Bernshtein, A. Flamholz, R. Phillips, and R. Milo. The total number and mass of SARS-CoV-2 virions. *Proceedings of the National Academy of Sciences*, 118(25), 2021. doi:10.1073/pnas.2024815118.

[52] O. Puhach, B. Meyer, and I. Eckerle. SARS-CoV-2 viral load and shedding kinetics. *Nature Reviews Microbiology*, 2023. doi:10.1038/s41579-022-00822-w.

[53] W. Sungnak, , N. Huang, C. Bécavin, M. Berg, R. Queen, M. Litvinukova, C. Talavera-López, H. Maatz, D. Reichart, F. Sampaziotis, K. B. Worlock, M. Yoshida, and J. L. Barnes. SARS-CoV-2 entry factors are highly expressed

in nasal epithelial cells together with innate immune genes. *Nature Medicine*, 26 (5):681–687, 2020. doi:10.1038/s41591-020-0868-6.

[54] M. Malik and T. Kunze. Detection of SARS-CoV-2 RNA in exhaled breath and its potential for prevention measures. *Infection Prevention in Practice*, 5(3):100299, 2023. doi:10.1016/j.infpip.2023.100299.

[55] J. Lai, K. K. Coleman, S. H. S. Tai, J. German, F. Hong, B. Albert, Y. Esparza, A. K. Srikakulapu, M. Schanz, I. S. Maldonado, M. Oertel, N. Fadul, T. L. Gold, S. Weston, K. Mullins, K. M. McPhaul, M. Frieman, and D. K. Milton. Exhaled Breath Aerosol Shedding of Highly Transmissible Versus Prior Severe Acute Respiratory Syndrome Coronavirus 2 Variants. *Clinical Infectious Diseases*, 76(5): 786–794, 2023. doi:10.1093/cid/ciac846.

[56] O. O. Adenaiye, J. Lai, P. J. B. de Mesquita, F. Hong, S. Youssefi, J. German, S. H. S. Tai, B. Albert, M. Schanz, S. Weston, J. Hang, C. Fung, H. K. Chung, K. K. Coleman, N. Sapoval, T. Treangen, I. M. Berry, K. Mullins, M. Frieman, T. Ma, and D. K. Milton. Infectious Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in Exhaled Aerosols and Efficacy of Masks During Early Mild Infection. *Clinical Infectious Diseases*, 75(1):e241–e248, 2022. doi:10.1093/cid/ciab797.

[57] M. Alsved, D. Nygren, S. Thuresson, P. Medstrand, C. J. Fraenkel, and J. Löndahl. SARS-CoV-2 in Exhaled Aerosol Particles from COVID-19 Cases and Its Association to Household Transmission. *Clinical Infectious Diseases*, 75(1):e50–e56, 2022. doi:10.1093/cid/ciac202.

[58] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, and W. Tan. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine*, 382(8):727–733, 2020. doi:10.1056/nejmoa2001017.

[59] L. Morawska, G. Buonanno, A. Mikszewski, and L. Stabile. The physics of respiratory particle generation, fate in the air, and inhalation. *Nature Reviews Physics*, 4(11):723–734, 2022. doi:10.1038/s42254-022-00506-7.

[60] A. Cvitešić Kušan, J. Baranašić, S. Frka, T. Lucijanić, A. Šribar, J. Knežević, G. Buonanno, and L. Stabile. The size distribution of SARS-CoV-2 genetic material in airborne particles sampled in hospital and home care environments occupied by COVID-19 positive subjects. *Science of The Total Environment*, 892:164642, 2023. doi:10.1016/j.scitotenv.2023.164642.

[61] R. B. Couch, T. R. Cate, R. G. Douglas, P. J. Gerone, and V. Knight. Effect of route of inoculation on experimental respiratory viral disease in volunteers and evidence for airborne transmission. *Bacteriological Reviews*, 30(3):517–529, 1966. doi:10.1128/br.30.3.517-529.1966.

[62] C. N. Haas, J. B. Rose, and C. P. Gerba. *Quantitative Microbial Risk Assessment*. John Wiley & Sons, Inc., Hoboken, NJ, 2nd edition, 2014. ISBN 9781118910023.

[63] D. K. Milton. What was the primary mode of smallpox transmission? Implications for biodefense. *Frontiers in Cellular and Infection Microbiology*, 2, 2012. doi:10.3389/fcimb.2012.00150.

[64] C. C. Wang, K. A. Prather, J. Sznitman, J. L. Jimenez, S. S. Lakdawala, Z. Tufekci, and L. C. Marr. Airborne transmission of respiratory viruses. *Science*, 373(6558): eabd9149, 2021. doi:10.1126/science.abd9149.

[65] Y. Drossinos, T. P. Weber, and N. I. Stilianakis. Droplets and aerosols: An artificial dichotomy in respiratory virus transmission. *Health Science Reports*, 4(2), 2021. doi:10.1002/hsr2.275.

[66] R. M. Jones and L. M. Brosseau. Aerosol Transmission of Infectious Disease. *Journal of Occupational and Environmental Medicine*, 57(5):501–508, 2015. doi:10.1097/jom.0000000000000448.

[67] R. Tellier, Y. Li, B. J. Cowling, and J. W. Tang. Recognition of aerosol transmission of infectious agents: a commentary. *BMC Infectious Diseases*, 19(1), 2019. doi:10.1186/s12879-019-3707-y.

[68] L. Morawska and D. K. Milton. It Is Time to Address Airborne Transmission of Coronavirus Disease 2019 (COVID-19). *Clinical Infectious Diseases*, 71(9):2311–2313, 2020. doi:10.1093/cid/ciaa939.

[69] R. Zhang, Y. Li, A. L. Zhang, Y. Wang, and M. J. Molina. Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proceedings of the National Academy of Sciences*, 117(26):14857–14863, 2020. doi:10.1073/pnas.2009637117.

[70] M. Bazant and J. Bush. A guideline to limit indoor airborne transmission of COVID-19. *Proceedings of the National Academy of Sciences*, 118(17), 2021. doi:10.1073/pnas.2018995118.

[71] S. L. Miller, W. W. Nazaroff, J. L. Jimenez, A. Boerstra, G. Buonanno, S. J. Dancer, J. Kurnitski, L. C. Marr, L. Morawska, and C. Noakes. Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the Skagit Valley Chorale superspreading event. *Indoor Air*, 31(2):314–323, 2020. doi:10.1111/ina.12751.

[72] G. Buonanno, L. Morawska, and L. Stabile. Quantitative assessment of the risk of airborne transmission of SARS-CoV-2 infection: Prospective and retrospective applications. *Environment International*, 145:106112, 2020. doi:10.1016/j.envint.2020.106112.

[73] J. Lu, J. Gu, K. Li, C. Xu, W. Su, Z. Lai, D. Zhou, C. Yu, B. Xu, and Z. Yang. COVID-19 Outbreak Associated with Air Conditioning in Restaurant, Guangzhou, China, 2020. *Emerging Infectious Diseases*, 26(7):1628–1631, 2020. doi:10.3201/eid2607.200764.

[74] Y. Li, H. Qian, J. Hang, X. Chen, P. Cheng, H. Ling, S. Wang, P. Liang, J. Li, S. Xiao, J. Wei, L. Liu, B. J. Cowling, and M. Kang. Probable airborne transmission of SARS-CoV-2 in a poorly ventilated restaurant. *Building and Environment*, 196: 107788, 2021. doi:10.1016/j.buildenv.2021.107788.

[75] D. Majra, J. Benson, J. Pitts, and J. Stebbing. SARS-CoV-2 (COVID-19) superspreader events. *Journal of Infection*, 82(1):36–40, 2021. doi:10.1016/j.jinf.2020.11.021.

[76] O. Wegehaupt, A. Endo, and A. Vassall. Superspreading, overdispersion and their implications in the SARS-CoV-2 (COVID-19) pandemic: a systematic review and meta-analysis of the literature. *BMC Public Health*, 23(1), 2023. doi:10.1186/s12889-023-15915-1.

[77] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066): 355–359, 2005. doi:10.1038/nature04153.

[78] M. E. J. Woolhouse, C. Dye, J.-F. Etard, T. Smith, J. D. Charlwood, G. P. Garnett, P. Hagan, J. L. K. Hii, P. D. Ndhlovu, R. J. Quinnell, C. H. Watts, S. K. Chandiwana, and R. M. Anderson. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proceedings of the National Academy of Sciences*, 94(1):338–342, 1997. doi:10.1073/pnas.94.1.338.

[79] N. van Doremalen, T. Bushmaker, D. H. Morris, A. Gamble, B. N. Williamson, A. Tamin, J. L. Harcourt, N. J. Thornburg, S. I. Gerber, J. O. Lloyd-Smith, E. de Wit, and V. J. Munste. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *New England Journal of Medicine*, 382(16):1564–1567, 2020. doi:10.1056/NEJMc2004973.

[80] G. Kampf, D. Todt, S. Pfaender, and E. Steinmann. Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. *Journal of Hospital Infection*, 104(3):246–251, 2020. doi:10.1016/j.jhin.2020.01.022.

[81] E. Goldman. Exaggerated risk of transmission of COVID-19 by fomites. *The Lancet Infectious Diseases*, 20(8):892–893, 2020. doi:10.1016/s1473-3099(20)30561-2.

[82] A. P. Harvey, E. R. Fuhrmeister, M. E. Cantrell, A. K. Pitol, J. M. Swarthout, J. E. Powers, M. L. Nadimpalli, T. R. Julian, and A. J. Pickering. Longitudinal Monitoring of SARS-CoV-2 RNA on High-Touch Surfaces in a Community Setting. *Environmental Science & Technology Letters*, 8(2):168–175, 2021. doi:10.1021/acs.estlett.0c00875.

[83] E. Goldman. SARS Wars: the Fomites Strike Back. *Applied and Environmental Microbiology*, 87(13), 2021. doi:10.1128/aem.00653-21.

[84] M. U. Mondelli, M. Colaneri, E. M. Seminari, F. Baldanti, and R. Bruno. Low risk of SARS-CoV-2 transmission by fomites in real-life conditions. *The Lancet Infectious Diseases*, 21(5):e112, 2021. doi:10.1016/s1473-3099(20)30678-2.

[85] M. D. Sobsey. Absence of virological and epidemiological evidence that SARS-CoV-2 poses COVID-19 risks from environmental fecal waste, wastewater and water exposures. *Journal of Water and Health*, 20(1):126–138, 2022. doi:10.2166/wh.2021.182.

[86] W. C. Hinds. *Aerosol Technology*. John Wiley & Sons, New York, NY, 2nd edition, 1999. ISBN 0471194107.

[87] O. B. Kudryashova, E. V. Muravlev, A. A. Antonnikova, and S. S. Titov. Propagation of viral bioaerosols indoors. *PLOS ONE*, 16(1):e0244983, 2021. doi:10.1371/journal.pone.0244983.

[88] V. Groma, S. Kugler, Á. Farkas, P. Füri, B. Madas, A. Nagy, T. Erdélyi, A. Horváth, V. Müller, R. Szántó-Egész, A. Micsinai, G. Gálffy, and J. Osán. Size distribution and relationship of airborne SARS-CoV-2 RNA to indoor aerosol in hospital ward environments. *Scientific Reports*, 13(1), 2023. doi:10.1038/s41598-023-30702-z.

[89] S. D. Judson and V. J. Munster. Nosocomial Transmission of Emerging Viruses via Aerosol-Generating Medical Procedures. *Viruses*, 11(10):940, 2019. doi:10.3390/v11100940.

[90] A. Davies, G. Thomson, J. Walker, and A. Bennett. A review of the risks and disease transmission associated with aerosol generating medical procedures. *Journal of Infection Prevention*, 10(4):122–126, 2009. doi:10.1177/1757177409106456.

[91] B. Haji Ali, M. S. Shahin, M. M. Masoumi Sangani, M. Faghihinezhad, and M. Baghdadi. Wastewater aerosols produced during flushing toilets, WWTPs, and irrigation with reclaimed municipal wastewater as indirect exposure to SARS-CoV-2. *Journal of Environmental Chemical Engineering*, 9(5):106201, 2021. doi:10.1016/j.jece.2021.106201.

[92] J. P. Duguid. The numbers and the sites of origin of the droplets expelled during expiratory activities. *Edinburgh Medical Journal*, 52(11):385–401, 1945. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5286249/`. Accessed on April 5, 2024.

[93] R. S. Papineni and F. S. Rosenthal. The Size Distribution of Droplets in the Exhaled Breath of Healthy Human Subjects. *Journal of Aerosol Medicine*, 10(2): 105–116, 1997. doi:10.1089/jam.1997.10.105.

[94] L. Morawska, G. R. Johnson, Z. D. Ristovski, M. Hargreaves, K. Mengersen, S. Corbett, C. Y. H. Chao, Y. Li, and D. Katoshevski. Size distribution and sites of origin of droplets expelled from the human respiratory tract during expiratory activities. *Journal of Aerosol Science*, 40(3):256–269, 2009. doi:10.1016/j.jaerosci.2008.11.002.

[95] G. Johnson, L. Morawska, Z. Ristovski, M. Hargreaves, K. Mengersen, C. Chao, M. Wan, Y. Li, X. Xie, D. Katoshevski, and S. Corbett. Modality of human expired aerosol size distributions. *Journal of Aerosol Science*, 42(12):839–851, 2011. doi:10.1016/j.jaerosci.2011.07.009.

[96] G. Zayas, M. C. Chiang, E. Wong, F. MacDonald, C. F. Lange, A. Senthilselvan, and M. King. Cough aerosol in healthy participants: fundamental knowledge to optimize droplet-spread infectious respiratory disease management. *BMC Pulmonary Medicine*, 12(1), 2012. doi:10.1186/1471-2466-12-11.

[97] G. Bagheri, O. Schlenczek, L. Turco, B. Thiede, K. Stieger, J. M. Kosub, S. Clauberg, M. L. Pöhlker, C. Pöhlker, J. Moláček, S. Scheithauer, and E. Bodenschatz. Size, concentration, and origin of human exhaled particles and their dependence on human factors with implications on infection transmission. *Journal of Aerosol Science*, 168:106102, 2023. doi:10.1016/j.jaerosci.2022.106102.

[98] M. L. Pöhlker, C. Pöhlker, O. O. Krüger, J.-D. Förster, T. Berkemeier, W. Elbert, J. Fröhlich-Nowoisky, U. Pöschl, G. Bagheri, E. Bodenschatz, J. A. Huffman, S. Scheithauer, and E. Mikhailov. Respiratory aerosols and droplets in the transmission of infectious diseases. *Reviews of Modern Physics*, 95(4):045001, 2023. doi:10.1103/revmodphys.95.045001.

[99] M. L. Pöhlker, O. O. Krüger, J.-D. Förster, T. Berkemeier, W. Elbert, J. Fröhlich-Nowoisky, U. Pöschl, C. Pöhlker, G. Bagheri, E. Bodenschatz, J. A. Huffman, S. Scheithauer, and E. Mikhailov. Respiratory aerosols and droplets in the transmission of infectious diseases. Online, 2021. doi:10.48550/ARXIV.2103.01188.

[100] G. Scheuch. Breathing Is Enough: For the Spread of Influenza Virus and SARS-CoV-2 by Breathing Only. *Journal of Aerosol Medicine and Pulmonary Drug Delivery*, 33(4):230–234, 2020. doi:10.1089/jamp.2020.1616.

[101] D. K. Milton, M. P. Fabian, B. J. Cowling, M. L. Grantham, and J. J. McDevitt. Influenza Virus Aerosols in Human Exhaled Breath: Particle Size, Culturability, and Effect of Surgical Masks. *PLoS Pathogens*, 9(3):e1003205, 2013. doi:10.1371/journal.ppat.1003205.

[102] M. Alsved, D. Nygren, S. Thuresson, C.-J. Fraenkel, P. Medstrand, and J. Löndahl. Size distribution of exhaled aerosol particles containing SARS-CoV-2 RNA. *Infectious Diseases*, 55(2):158–163, 2023. doi:10.1080/23744235.2022.2140822.

[103] C. S. Ng, K. L. Chong, R. Yang, M. Li, R. Verzicco, and D. Lohse. Growth of respiratory droplets in cold and humid air. *Physical Review Fluids*, 6(5):054303, 2021. doi:10.1103/PhysRevFluids.6.054303.

[104] K. L. Chong, C. S. Ng, N. Hori, R. Yang, R. Verzicco, and D. Lohse. Extended Lifetime of Respiratory Droplets in a Turbulent Vapor Puff and Its Implications on Airborne Disease Transmission. *Physical Review Letters*, 126(3):034502, 2021. doi:10.1103/physrevlett.126.034502.

[105] X. Xie, Y. Li, A. T. Y. Chwang, P. L. Ho, and W. H. Seto. How far droplets can move in indoor environments – revisiting the Wells evaporation-falling curve. *Indoor Air*, 17(3):211–225, 2007. doi:10.1111/j.1600-0668.2007.00469.x.

[106] K. W. Mui, L. T. Wong, C. L. Wu, and A. C. K. Lai. Numerical modeling of exhaled droplet nuclei dispersion and mixing in indoor environments. *Journal of Hazardous Materials*, 167(1–3):736–744, 2009. doi:10.1016/j.jhazmat.2009.01.041.

[107] M. Hossain and N. H. Faisal. Modeling aerosol cloud aerodynamics during human coughing, talking, and breathing actions. *AIP Advances*, 11(4):045111, 2021. doi:10.1063/5.0042952.

[108] Y. Ji, H. Qian, J. Ye, and X. Zheng. The impact of ambient humidity on the evaporation and dispersion of exhaled breathing droplets: A numerical investigation. *Journal of Aerosol Science*, 115:164–172, 2018. doi:10.1016/j.jaerosci.2017.10.009.

[109] S. Chatterjee, J. S. Murallidharan, A. Agrawal, and R. Bhardwaj. How coronavirus survives for hours in aerosols. *Physics of Fluids*, 33(8):081708, 2021. doi:10.1063/5.0059908.

[110] S. J. Smither, L. S. Eastaugh, J. S. Findlay, and M. S. Lever. Experimental aerosol survival of SARS-CoV-2 in artificial saliva and tissue culture media at medium and high humidity. *Emerging Microbes & Infections*, 9(1):1415–1417, 2020. doi:10.1080/22221751.2020.1777906.

[111] S. H. Smith, G. A. Somsen, C. van Rijn, S. Kooij, L. van der Hoek, R. A. Bem, and D. Bonn. Aerosol persistence in relation to possible transmission of SARS-CoV-2. *Physics of Fluids*, 32(10):107108, 2020. doi:10.1063/5.0027844.

[112] G. Somsen, C. van Rijn, S. Kooij, R. Bem, and D. Bonn. Small droplet aerosols in poorly ventilated spaces and SARS-CoV-2 transmission. *The Lancet Respiratory Medicine*, 8(7):658–659, 2020. doi:10.1016/s2213-2600(20)30245-9.

[113] International Commission on Radiological Protection. Human Respiratory Tract Model for Radiological Protection. *Annals of the ICRP*, 24(1–3), 1994. URL `https://journals.sagepub.com/doi/pdf/10.1177/ANIB_24_1-3`. Accessed on April 5, 2024.

[114] T. R. Sosnowski. Inhaled aerosols: Their role in COVID-19 transmission, including biophysical interactions in the lungs. *Current Opinion in Colloid & Interface Science*, 54:101451, 2021. doi:10.1016/j.cocis.2021.101451.

[115] H. Li, F. Y. Leong, G. Xu, C. W. Kang, K. H. Lim, B. H. Tan, and C. M. Loo. Airborne dispersion of droplets during coughing: a physical model of viral transmission. *Scientific Reports*, 11(1), 2021. doi:10.1038/s41598-021-84245-2.

[116] I. Echeverría-Huarte, A. Garcimartín, R. C. Hidalgo, C. Martín-Gómez, and I. Zuriguel. Estimating density limits for walking pedestrians keeping a safe interpersonal distancing. *Scientific Reports*, 11(1), 2021. doi:10.1038/s41598-020-79454-0.

[117] L. Boulos, J. A. Curran, A. Gallant, H. Wong, C. Johnson, A. Delahunty-Pike, L. Saxinger, D. Chu, J. Comeau, T. Flynn, J. Clegg, and C. Dye. Effectiveness of face masks for reducing transmission of SARS-CoV-2: a rapid systematic review. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 381(2257), 2023. doi:10.1098/rsta.2023.0133.

[118] Deutsche Telekom AG and SAP SE. Corona-Warn-App. Online, 2023. URL `https://github.com/corona-warn-app/`. Accessed on April 5, 2024.

[119] A. J. Kucharski, P. Klepac, A. J. K. Conlan, S. M. Kissler, M. L. Tang, H. Fry, J. R. Gog, W. J. Edmunds, J. C. Emery, G. Medley, J. D. Munday, T. W. Russell, Q. J. Leclerc, C. Diamond, S. R. Procter, A. Gimma, F. Y. Sun, H. P. Gibbs, A. Rosello, K. van Zandvoort, S. Hué, S. R. Meakin, A. K. Deol, G. Knight, T. Jombart, A. M. Foss, N. I. Bosse, K. E. Atkins, B. J. Quilty, R. Lowe, K. Prem, S. Flasche, C. A. B. Pearson, R. M. G. J. Houben, E. S. Nightingale, A. Endo, D. C. Tully, Y. Liu, J. Villabona-Arenas, K. O'Reilly, S. Funk, R. M. Eggo, M. Jit, E. M. Rees, J. Hellewell, S. Clifford, C. I. Jarvis, S. Abbott, M. Auzenbergs, N. G. Davies, and D. Simons. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *The Lancet Infectious Diseases*, 20(10):1151–1160, 2020. doi:10.1016/s1473-3099(20)30457-6.

[120] M. E. Kretzschmar, G. Rozhnova, M. C. J. Bootsma, M. van Boven, J. H. H. M. van de Wijgert, and M. J. M. Bonten. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *The Lancet Public Health*, 5 (8):e452–e459, 2020. doi:10.1016/s2468-2667(20)30157-2.

[121] A. Mendez-Brito, C. El Bcheraoui, and F. Pozo-Martin. Systematic review of empirical studies comparing the effectiveness of non-pharmaceutical interventions against COVID-19. *Journal of Infection*, 83(3):281–293, 2021. doi:10.1016/j.jinf.2021.06.018.

[122] J. P. A. Ioannidis, M. Salholz-Hillel, K. W. Boyack, and J. Baas. The rapid, massive growth of COVID-19 authors in the scientific literature. *Royal Society Open Science*, 8(9), 2021. doi:10.1098/rsos.210389.

[123] F. Krammer. SARS-CoV-2 vaccines in development. *Nature*, 586(7830):516–527, 2020. doi:10.1038/s41586-020-2798-3.

[124] S. Gianola, T. S. Jesus, S. Bargeri, and G. Castellini. Characteristics of academic publications, preprints, and registered clinical trials on the COVID-19 pandemic. *PLOS ONE*, 15(10):e0240123, 2020. doi:10.1371/journal.pone.0240123.

[125] N. Fraser, L. Brierley, G. Dey, J. K. Polka, M. Pálfy, F. Nanni, and J. A. Coates. The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLOS Biology*, 19(4):e3000959, 2021. doi:10.1371/journal.pbio.3000959.

[126] D. Lewis. Superspreading drives the COVID pandemic — and could help to tame it. *Nature*, 590(7847):544–546, 2021. doi:10.1038/d41586-021-00460-x.

[127] F. Wong and J. J. Collins. Evidence that coronavirus superspreading is fat-tailed. *Proceedings of the National Academy of Sciences*, 117(47):29416–29418, 2020. doi:10.1073/pnas.2018490117.

[128] X. Yu, D. Ran, J. Wang, Y. Qin, R. Liu, X. Shi, Y. Wang, C. Xie, J. Jiang, and J. Zhou. Unclear but present danger: An asymptomatic SARS-CoV-2 carrier. *Genes & Diseases*, 7(4):558–566, 2020. doi:10.1016/j.gendis.2020.07.010.

[129] L. Hamner, P. Dubbel, I. Capron, A. Ross, A. Jordan, J. Lee, J. Lynn, A. Ball, S. Narwal, S. Russell, D. Patrick, and H. Leibrand. High SARS-CoV-2 Attack Rate Following Exposure at a Choir Practice — Skagit County, Washington, March 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69(19):606–610, 2020. doi:10.15585/mmwr.mm6919e6.

[130] A. R. McLean. *Infectious Disease Modeling*, pages 5347–5357. Springer New York, New York, NY, 2012. doi:10.1007/978-1-4419-0851-3_539.

[131] W. H. Hamer. The Milroy Lectures on Epidemic Disease in England - The Evidence of Variability and of Persistency of Type. *The Lancet*, 167(4307):733–739, 1906. doi:10.1016/s0140-6736(01)80340-8. Originally published as Volume 1, Issue 4307.

[132] R. Ross. Some a priori pathometric equations. *BMJ*, 1(2830):546–547, 1915. doi:10.1136/bmj.1.2830.546.

[133] R. Ross and H. P. Hudson. An application of the theory of probabilities to the study of a priori pathometry. - Part III. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 93(650): 225–240, 1917. doi:10.1098/rspa.1917.0015.

[134] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, 1927. doi:10.1098/rspa.1927.0118.

[135] T. Harko, F. S. N. Lobo, and M. K. Mak. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Applied Mathematics and Computation*, 236:184–194, 2014. doi:10.1016/j.amc.2014.03.030.

[136] T. Lazebnik. Computational applications of extended SIR models: A review focused on airborne pandemics. *Ecological Modelling*, 483:110422, 2023. doi:10.1016/j.ecolmodel.2023.110422.

[137] H. W. Hethcote. The Mathematics of Infectious Diseases. *SIAM Review*, 42(4): 599–653, 2000. doi:10.1137/s0036144500371907.

[138] M. E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66 (1):016128, 2002. doi:10.1103/physreve.66.016128.

[139] M. J. Keeling and K. T. D. Eames. Networks and epidemic models. *Journal of The Royal Society Interface*, 2(4):295–307, 2005. doi:10.1098/rsif.2005.0051.

[140] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925–979, 2015. doi:10.1103/revmodphys.87.925.

[141] M. E. Halloran, I. M. Longini, A. Nizam, and Y. Yang. Containing Bioterrorist Smallpox. *Science*, 298(5597):1428–1432, 2002. doi:10.1126/science.1074674.

[142] S. M. Jenness, K. S. Willebrand, A. A. Malik, B. A. Lopman, and S. B. Omer. Dynamic network strategies for SARS-CoV-2 control on a cruise ship. *Epidemics*, 37:100488, 2021. doi:10.1016/j.epidem.2021.100488.

[143] X. Guo, A. Gupta, A. Sampat, and C. Zhai. A stochastic contact network model for assessing outbreak risk of COVID-19 in workplaces. *PLOS ONE*, 17(1):e0262316, 2022. doi:10.1371/journal.pone.0262316.

[144] L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences*, 101(42): 15124–15129, 2004. doi:10.1073/pnas.0308344101.

[145] S. Eubank, H. Guclu, V. S. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004. doi:10.1038/nature02541.

[146] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, and A. Vespignani. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489):395–400, 2020. doi:10.1126/science.aba9757.

[147] S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky, and J. Leskovec. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589(7840):82–87, 2021. doi:10.1038/s41586-020-2923-3.

[148] M. J. Kühn, D. Abele, T. Mitra, W. Koslow, M. Abedi, K. Rack, M. Siggel, S. Khailaie, M. Klitz, S. Binder, L. Spataro, J. Gilg, J. Kleinert, M. Häberle, L. Plötzke, C. D. Spinner, M. Stecher, X. X. Zhu, A. Basermann, and M. 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*, 339:108648, 2021. doi:10.1016/j.mbs.2021.108648.

[149] S. A. Müller, M. Balmer, W. Charlton, R. Ewert, A. Neumann, C. Rakow, T. Schlenther, and K. Nagel. Predicting the effects of COVID-19 related interventions in urban settings by combining activity-based modelling, agent-based simulation, and mobile phone data. *PLOS ONE*, 16(10):e0259037, 2021. doi:10.1371/journal.pone.0259037.

[150] M. J. Kühn, D. Abele, D. Kerkmann, S. Korf, H. Zunker, A. Wendler, J. Bicker, K. Nguyen, R. Schmieding, L. Plötzke, P. Lenz, M. Betz, C. Gerstein, A. Schmidt, P. Johannssen, M. Klitz, W. Koslow, S. Binder, M. Siggel, J. Kleinert, K. Rack, A. Lutz, and M. Meyer-Hermann. MEmilio v1.0.0 - A high performance Modular EpideMIcs simuLatIOn software. online. doi:10.5281/ZENODO.10412635.

[151] K. W. Axhausen. *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press, London, 2016. doi:10.5334/baw.

[152] C. J. Noakes, C. B. Beggs, P. A. Sleigh, and K. G. Kerr. Modelling the transmission of airborne infections in enclosed spaces. *Epidemiology and Infection*, 134(5):1082–1091, 2006. doi:10.1017/s0950268806005875.

[153] C. B. Beggs, C. J. Noakes, P. A. Sleigh, L. A. Fletcher, and K. Siddiqi. The transmission of tuberculosis in confined spaces: an analytical review of alternative epidemiological models. *The International Journal of Tuberculosis and Lung Disease*, 7 (11):1015–1026, 2003. ISSN 1027-3719. URL `https://www.ingentaconnect.com/content/iuatld/ijtld/2003/00000007/00000011/art00002`. Accessed on October 10, 2023.

[154] L. Gammaitoni. Using a Mathematical Model to Evaluate the Efficacy of TB Control Measures. *Emerging Infectious Diseases*, 3(3):335–342, 1997. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2627642/`. Accessed on April 5, 2024.

[155] M. Nicas. An Analytical Framework for Relating Dose, Risk, and Incidence: An Application to Occupational Tuberculosis Infection. *Risk Analysis*, 16(4):527–538, 1996. doi:10.1111/j.1539-6924.1996.tb01098.x.

[156] W. W. Nazaroff, M. Nicas, and S. L. Miller. Framework for Evaluating Measures to Control Nosocomial Tuberculosis Transmission. *Indoor Air*, 8(4):205–218, 1998. doi:10.1111/j.1600-0668.1998.00002.x.

[157] G. Buonanno, L. Stabile, and L. Morawska. Estimation of airborne viral emission: Quanta emission rate of SARS-CoV-2 for infection risk assessment. *Environment International*, 141:105794, 2020. doi:10.1016/j.envint.2020.105794.

[158] J. S. Salinas, K. A. Krishnaprasad, N. Zgheib, and S. Balachandar. Improved guidelines of indoor airborne transmission taking into account departure from the well-mixed assumption. *Physical Review Fluids*, 7(6):064309, 2022. doi:10.1103/physrevfluids.7.064309.

[159] Z. Lau, I. M. Griffiths, A. English, and K. Kaouri. Predicting the spatio-temporal infection risk in indoor spaces using an efficient airborne transmission model. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 478(2259), 2022. doi:10.1098/rspa.2021.0383.

[160] I. Martinez, J. L. Bruse, A. M. Florez-Tapia, E. Viles, and I. G. Olaizola. ArchABM: An agent-based simulator of human interaction with the built environment. CO2 and viral load analysis for indoor air quality. *Building and Environment*, 207: 108495, 2022. doi:10.1016/j.buildenv.2021.108495.

[161] M. Gaddis, V. Manoranjan, and J. Streipel. Agent-Based Modeling to Simulate Aerosolized Transmission of SARS-CoV-2 inside Small Ventilated Spaces. *COVID*, 3(7):937–955, 2023. doi:10.3390/covid3070068.

[162] Y. Yan, X. Li, X. Fang, Y. Tao, and J. Tu. A spatiotemporal assessment of occupants' infection risks in a multi-occupants space using modified Wells-Riley model. *Building and Environment*, 230:110007, 2023. doi:10.1016/j.buildenv.2023.110007.

[163] A. Booth, A. Sutton, and D. Papaioannou. *Systematic approaches to a successful literature review*. SAGE, 2nd edition, 2016. ISBN 9781473912465.

[164] A. Templeton, H. Xie, S. Gwynne, A. Hunt, P. Thompson, and G. Köster. Agent-based models of social behaviour and communication in evacuations: A systematic review. *Safety Science*, 176:106520, 2024. doi:10.1016/j.ssci.2024.106520.

[165] A. Johansson, M. Batty, K. Hayashi, O. A. Bar, D. Marcozzi, and Z. Memish. Crowd and environmental management during mass gatherings. *The Lancet Infectious Diseases*, 12(2):150–156, 2012. doi:10.1016/s1473-3099(11)70287-0.

[166] L. Goscé, D. Barton, and A. Johansson. Analytical Modelling of the Spread of Disease in Confined and Crowded Spaces. *Scientific Reports*, 4(1), 2014. doi:10.1038/srep04856.

[167] S. Namilae, A. Srinivasan, A. Mubayi, M. Scotch, and R. Pahle. Self-propelled pedestrian dynamics model: Application to passenger movement and infection propagation in airplanes. *Physica A: Statistical Mechanics and its Applications*, 465:248–260, 2017. doi:10.1016/j.physa.2016.08.028.

[168] S. Namilae, P. Derjany, A. Mubayi, M. Scotch, and A. Srinivasan. Multiscale model for pedestrian and infection dynamics during air travel. *Physical Review E*, 95(5): 052320, 2017. doi:10.1103/physreve.95.052320.

[169] P. Derjany and S. Namilae. Computational Model for Pedestrian Movement and Infectious Diseases Spread During Air Travel. In *2018 AIAA Modeling and Simulation Technologies Conference*. American Institute of Aeronautics and Astronautics, 2018. doi:10.2514/6.2018-0419.

[170] S. Namilae, P. Derjany, D. Liu, A. Mubayi, and A. Srinivasan. Multiscale Pedestrian Dynamics and Infection Spread Model for Policy Analysis. In *Proceedings from the 9th International Conference on Pedestrian and Evacuation Dynamics (PED2018) Lund, Sweden - August 21–23, 2018*, volume 5, pages 512–514. Forschungszentrum Jülich, Zentralbibliothek, 2020. doi:10.17815/cd.2020.86.

[171] P. Derjany, S. Namilae, D. Liu, and A. Srinivasan. Computational Modeling Framework for the Study of Infectious Disease Spread through Commercial Air-Travel. In *2020 IEEE Aerospace Conference*. IEEE, 2020. doi:10.1109/aero47225.2020.9172285.

[172] P. Derjany, S. Namilae, D. Liu, and A. Srinivasan. Multiscale model for the optimal design of pedestrian queues to mitigate infectious disease spread. *PLOS ONE*, 15 (7):e0235891, 2020. doi:10.1371/journal.pone.0235891.

[173] M. Abadeer and S. Gorlatch. Simulating infection transmission: A case study of COVID-19. In A. Nketsa, C. Baron, and C. Foucher, editors, *The European Simulation and Modelling Conference, ESM 2020*, pages 310–317. EUROSIS, 2020. URL `https://www.scopus.com/inward/record.uri?eid=2-s2.0-85096745031&partnerID=40&md5=fd5dee05ef393a85cde40b5d42e55ef4`. Accessed on April 5, 2024.

[174] M. Abadeer, S. Magharious, and S. Gorlatch. Modeling and interactive simulation of measures against infection transmission. *SIMULATION*, page 003754972211338, 2022. doi:10.1177/00375497221133849.

[175] K. Rathinakumar and A. Quaini. A microscopic approach to study the onset of a highly infectious disease spreading. *Mathematical Biosciences*, 329:108475, 2020. doi:10.1016/j.mbs.2020.108475.

[176] Q. Xu and M. Chraibi. On the Effectiveness of the Measures in Supermarkets for Reducing Contact among Customers during COVID-19 Period. *Sustainability*, 12 (22):9385, 2020. doi:10.3390/su12229385.

[177] S. Butail and M. Porfiri. The Effect of An Emergency Evacuation on the Spread of COVID19. *Frontiers in Physics*, 8, 2021. doi:10.3389/fphy.2020.631264.

[178] M. D'Orazio, G. Bernardini, and E. Quagliarini. A probabilistic model to evaluate the effectiveness of main solutions to COVID-19 spreading in university buildings according to proximity and time-based consolidated criteria. *Building Simulation*, 14(6):1795–1809, 2021. doi:10.1007/s12273-021-0770-2.

[179] M. D'Orazio, G. Bernardini, and E. Quagliarini. Sustainable and resilient strategies for touristic cities against COVID-19: An agent-based approach. *Safety Science*, 142:105399, 2021. doi:10.1016/j.ssci.2021.105399.

[180] W. Garcia, S. Mendez, B. Fray, and A. Nicolas. Model-based assessment of the risks of viral transmission in non-confined crowds. *Safety Science*, 144:105453, 2021. doi:10.1016/j.ssci.2021.105453.

[181] F. Ying and N. O'Clery. Modelling COVID-19 transmission in supermarkets using an agent-based model. *PLOS ONE*, 16(4):e0249821, 2021. doi:10.1371/journal.pone.0249821.

[182] Y. Xiao, M. Yang, Z. Zhu, H. Yang, L. Zhang, and S. Ghader. Modeling indoor-level non-pharmaceutical interventions during the COVID-19 pandemic: A pedestrian dynamics-based microscopic simulation approach. *Transport Policy*, 109:12–23, 2021. doi:10.1016/j.tranpol.2021.05.004.

[183] H. Cui, J. Xie, M. Zhu, X. Tian, and C. Wan. Virus transmission risk of college students in railway station during Post-COVID-19 era: Combining the social force model and the virus transmission model. *Physica A: Statistical Mechanics and its Applications*, 608:128284, 2022. doi:10.1016/j.physa.2022.128284.

[184] Z. Cui, M. Cai, Y. Xiao, Z. Zhu, M. Yang, and G. Chen. Forecasting the transmission trends of respiratory infectious diseases with an exposure-risk-based model at the microscopic level. *Environmental Research*, 212:113428, 2022. doi:10.1016/j.envres.2022.113428.

[185] S. Namilae, Y. Wu, A. Mubayi, A. Srinivasan, and M. Scotch. Identifying mitigation strategies for COVID-19 superspreading on flights using models that account for passenger movement. *Travel Medicine and Infectious Disease*, 47:102313, 2022. doi:10.1016/j.tmaid.2022.102313.

[186] Y. Wu, S. Namilae, A. Mubayi, M. Scotch, and A. Srinivasan. Incorporating Pedestrian Movement in Computational Models of COVID-19 Spread during Air-travel. In *2022 IEEE Aerospace Conference (AERO)*. IEEE, 2022. doi:10.1109/aero53065.2022.9843497.

[187] A. Rodríguez, E. Cuevas, D. Zaldivar, B. Morales-Castañeda, R. Sarkar, and E. Houssein. An agent-based transmission model of COVID-19 for reopening policy design. *Computers in Biology and Medicine*, 148:105847, 2022. doi:10.1016/j.compbiomed.2022.105847.

[188] H. Zhu, Y. Qin, Q. Zhao, and Q. Zhao. A Hybrid Model for the Impact of COVID-19 Prevention Measures on the Sustainable Development of the Aviation System. *Mathematical Problems in Engineering*, 2022:1–15, 2022. doi:10.1155/2022/5430328.

[189] T. Harweg, D. Bachmann, and F. Weichert. Agent-based simulation of pedestrian dynamics for exposure time estimation in epidemic risk assessment. *Journal of Public Health*, 31(2):221–228, 2023. doi:10.1007/s10389-021-01489-y.

[190] H. Karimian, Q. Fan, Q. Li, Y. Chen, and J. Shi. Spatiotemporal transmission of infectious particles in environment: A case study of Covid-19. *Chemosphere*, 335: 139065, 2023. doi:10.1016/j.chemosphere.2023.139065.

[191] S. Mendez, W. Garcia, and A. Nicolas. From Microscopic Droplets to Macroscopic Crowds: Crossing the Scales in Models of Short-Range Respiratory Disease Transmission, with Application to COVID-19. *Advanced Science*, 2023. doi:10.1002/advs.202205255.

[192] Q. Qiao, C. Cheung, A. Yunusa-Kaltungo, P. Manu, R. Cao, and Z. Yuan. An interactive agent-based modelling framework for assessing COVID-19 transmission risk on construction site. *Safety Science*, 168:106312, 2023. doi:10.1016/j.ssci.2023.106312.

[193] A. Nicolas and S. Mendez. Viral transmission in pedestrian crowds: Coupling an open-source code assessing the risks of airborne contagion with diverse pedestrian dynamics models. Online, 2023. doi:10.48550/arXiv.2312.01779.

[194] D. Duives, Y. Chang, M. Sparnaaij, B. Wouda, D. Boschma, Y. Liu, Y. Yuan, W. Daamen, M. de Jong, C. Teberg, K. Schachtschneider, R. Sikkema, L. van Veen, and Q. ten Bosch. The multi-dimensional challenges of controlling SARS-CoV-2 transmission in indoor spaces: Insights from the linkage of a microscopic pedestrian simulation and virus transmission models, 2021. doi:10.1101/2021.04.12.21255349.

[195] R. Löhner and H. Antil. High fidelity modeling of aerosol pathogen propagation in built environments with moving pedestrians. *International Journal for Numerical Methods in Biomedical Engineering*, 37(3), 2021. doi:10.1002/cnm.3428.

[196] R. Löhner, H. Antil, A. Srinivasan, S. Idelsohn, and E. Oñate. High-Fidelity Simulation of Pathogen Propagation, Transmission and Mitigation in the Built Environment. *Archives of Computational Methods in Engineering*, 28(6):4237–4262, 2021. doi:10.1007/s11831-021-09606-6.

[197] R. Löhner, H. Antil, J. M. Gimenez, S. Idelsohn, and E. Oñate. A deterministic pathogen transmission model based on high-fidelity physics. *Computer Methods in Applied Mechanics and Engineering*, 401:114929, 2022. doi:10.1016/j.cma.2022.114929.

[198] M. Choi and A. Hohl. Investigating factors in indoor transmission of respiratory disease through agent-based modeling. *Transactions in GIS*, 27(6):1794–1827, 2023. doi:10.1111/tgis.13099.

[199] D. S. I. Kanté, A. Jebrane, A. Bouchnita, and A. Hakim. Estimating the Risk of Contracting COVID-19 in Different Settings Using a Multiscale Transmission Dynamics Model. *Mathematics*, 11(1):254, 2023. doi:10.3390/math11010254.

[200] A. Bouchnita and A. Jebrane. A hybrid multi-scale model of COVID-19 transmission dynamics to assess the potential of non-pharmaceutical interventions. *Chaos, Solitons & Fractals*, 138:109941, 2020. doi:10.1016/j.chaos.2020.109941.

[201] E. Lanzarotti, L. Santi, R. Castro, F. Roslan, and L. Groisman. A Multi-Aspect Agent-Based Model of COVID-19: Disease Dynamics, Contact Tracing Interventions and Shared Space-Driven Contagions. In *2021 Winter Simulation Conference (WSC)*. IEEE, 2021. doi:10.1109/wsc52266.2021.9715445.

[202] S. Sajjadi, A. Hashemi, and F. Ghanbarnejad. Social distancing in pedestrian dynamics and its effect on disease spreading. *Physical Review E*, 104(1):014313, 2021. doi:10.1103/physreve.104.014313.

[203] Z. Han, S. Ma, C. Gao, E. Shao, Y. Xie, Y. Zhang, L. Geng, and Y. Li. Disease Simulation in Airport Scenario Based on Individual Mobility Model. *ACM Transactions on Intelligent Systems and Technology*, 2023. doi:10.1145/3593589. Just Accepted.

[204] E. Ronchi and R. Lovreglio. EXPOSED: An occupant exposure model for confined spaces to retrofit crowd models during a pandemic. *Safety Science*, 130(104834), 2020. doi:10.1016/j.ssci.2020.104834.

[205] L.-A. Cotfas, C. Delcea, R. J. Milne, and M. Salari. Evaluating Classical Airplane Boarding Methods Considering COVID-19 Flying Restrictions. *Symmetry*, 12(7): 1087, 2020. doi:10.3390/sym12071087.

[206] R. J. Milne, C. Delcea, L.-A. Cotfas, and C. Ioanas. Evaluation of Boarding Methods Adapted for Social Distancing When Using Apron Buses. *IEEE Access*, 8: 151650–151667, 2020. doi:10.1109/access.2020.3015736.

[207] R. J. Milne, L.-A. Cotfas, C. Delcea, L. Crăciun, and A.-G. Molănescu. Adapting the reverse pyramid airplane boarding method for social distancing in times of COVID-19. *PLOS ONE*, 15(11):e0242131, 2020. doi:10.1371/journal.pone.0242131.

[208] R. J. Milne, C. Delcea, and L.-A. Cotfas. Airplane boarding methods that reduce risk from COVID-19. *Safety Science*, 134:105061, 2021. doi:10.1016/j.ssci.2020.105061.

[209] A. Nicolas and S. Mendez. Viral Transmission in Pedestrian Crowds: Coupling an Open-source Code Assessing the Risks of Airborne Contagion with Diverse Pedestrian Dynamics Models. *Collective Dynamics*, 9:1–10, 2024. doi:10.17815/cd.2024.159.

[210] G. Visconti and P. Ruggieri. *Fluid Dynamics: Fundamentals and Applications*. Springer International Publishing, 1st edition, 2020. doi:10.1007/978-3-030-49562-6.

[211] B. Gorbunov. Aerosol Particles Generated by Coughing and Sneezing of a SARS-CoV-2 (COVID-19) Host Travel over 30 m Distance. *Aerosol and Air Quality Research*, 21(3):200468, 2021. doi:10.4209/aaqr.200468.

[212] S. Rahn and G. Köster. Reusable software structures for coupling agent-based locomotion models and disease transmission models. In K. R. Rao, A. Seyfried, and A. Schadschneider, editors, *Traffic and Granular Flow '22*, pages 239–246. Springer Nature Singapore, 2024. doi:10.1007/978-981-99-7976-9_30.

[213] S. Rahn, G. Köster, and H.-J. Bungartz. Toward unraveling airborne pathogen transmission in crowds: Parameter study for an agent-based exposure model. *Safety Science*, 175:106524, 2024. doi:10.1016/j.ssci.2024.106524.

[214] J. H. Ferziger and M. Perić. *Computational Methods for Fluid Dynamics*. Springer Berlin Heidelberg, 3rd edition, 2002. doi:10.1007/978-3-642-56026-2.

[215] T. Korhonen, S. Hostikka, S. Heliövaara, H. Ehtamo, and K. J. Matikainen. Integration of an agent based evacuation simulation and state-of-the-art fire simulation. In *Proceedings of the 7th Asia-Oceania Symposium on Fire Science & Technology, Hong Kong*, 2007. URL `https://pdfs.semanticscholar.org/5ea1/8b163221a6e736a47a0df9e785efbcf6df64.pdf`. Accessed on April 5, 2024.

[216] M. Benhassine, R. De Rouck, E. Dhondt, F. Van Utterbeeck, M. Debacker, and I. Hubloue. Simulating the Evacuation of a Subway Station after a Sarin Release. In *The 36th annual European Simulation and Modelling Conference*, 2023. URL `https://researchportal.vub.be/en/publications/simulating-the-evacuation-of-a-subway-station-after-a-sarin-relea`. Accessed on April 5, 2024.

[217] A. A. Sedighi, F. Haghighat, F. Nasiri, S. Cao, and C. Ren. Approaches in CFD modeling of respiratory droplet dispersion - issues and challenges. *Sustainable Cities and Society*, 97:104696, 2023. doi:10.1016/j.scs.2023.104696.

[218] Y. Sheikhnejad, R. Aghamolaei, M. Fallahpour, H. Motamedi, M. Moshfeghi, P. A. Mirzaei, and H. Bordbar. Airborne and aerosol pathogen transmission modeling of respiratory events in buildings: An overview of computational fluid dynamics. *Sustainable Cities and Society*, 79:103704, 2022. doi:10.1016/j.scs.2022.103704.

[219] A. Laitinen, M. Korhonen, K. Keskinen, O. Kaario, and V. Vuorinen. Large-eddy simulation of buoyant airflow in an airborne pathogen transmission scenario. *Building and Environment*, 241:110462, 2023. doi:10.1016/j.buildenv.2023.110462.

[220] B. Kleinmeier, B. Zönnchen, M. Gödel, and G. Köster. Vadere: An Open-Source Simulation Framework to Promote Interdisciplinary Understanding. *Collective Dynamics*, 4, 2019. doi:10.17815/CD.2019.21.

[221] M. Chraibi, K. Kratz, and T. Schrödter. JuPedSim, 2023. doi:10.5281/zenodo.10124693.

[222] A. Schadschneider, M. Chraibi, A. Seyfried, A. Tordeux, and J. Zhang. Pedestrian dynamics: From empirical results to modeling. *Modeling and Simulation in Science, Engineering and Technology*, pages 63–102, 2018. doi:10.1007/978-3-030-05129-7_4.

[223] B. Kleinmeier. *Modeling of Behavioral Changes in Agent-Based Simulations*. Dissertation, Technische Universität München, 2021. URL `http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20210521-1595400-1-2`. Accessed on April 5, 2024.

[224] A. Corbetta and F. Toschi. Physics of Human Crowds. *Annual Review of Condensed Matter Physics*, 14(1):311–333, 2023. doi:10.1146/annurev-conmatphys-031620-100450.

[225] M. Haghani. Crowd dynamics research in the era of Covid-19 pandemic: Challenges and opportunities. *Safety Science*, 153:105818, 2022. doi:10.1016/j.ssci.2022.105818.

[226] S. P. Hoogendoorn and P. H. L. Bovy. Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological*, 38(2): 169–190, 2004. doi:10.1016/S0191-2615(03)00007-9.

[227] K. Hirai and K. Tarui. A simulation of the behavior of a crowd in panic. In *Proc. of the 1975 International Conference on Cybernetics and Society*, page 409, 1975.

[228] D. Helbing and P. Molnár. Social Force Model for pedestrian dynamics. *Physical Review E*, 51(5):4282–4286, 1995. doi:10.1103/PhysRevE.51.4282.

[229] A. Johansson, D. Helbing, and P. Shukla. Specification of the social force pedestrian model by evolutionary adjustment to video tracking data. *Advances in Complex Systems*, 10:271–288, 2007. doi:10.1142/S0219525907001355.

[230] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz. The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics. *PLoS ONE*, 5(4):e10047, 2010. doi:10.1371/journal.pone.0010047.

[231] M. Chraibi, A. Seyfried, and A. Schadschneider. Generalized centrifugal-force model for pedestrian dynamics. *Physical Review E*, 82(4):046111, 2010. doi:10.1103/PhysRevE.82.046111.

[232] G. Köster, F. Treml, and M. Gödel. Avoiding numerical pitfalls in social force models. *Physical Review E*, 87(6):063305, 2013. doi:10.1103/PhysRevE.87.063305.

[233] W. J. Yu, R. Chen, L. Y. Dong, and S. Q. Dai. Centrifugal force model for pedestrian dynamics. *Physical Review E*, 72:026112, 2005. doi:10.1103/PhysRevE.72.026112.

[234] T. I. Lakoba, D. J. Kaup, and N. M. Finkelstein. Modifications of the Helbing-Molnár-Farkas-Vicsek Social Force Model for Pedestrian Evolution. *Simulation*, 81 (5):339–352, 2005. doi:10.1177/0037549705052772.

[235] D. R. Parisi, M. Gilman, and H. Moldovan. A modification of the Social Force Model can reproduce experimental data of pedestrian flows in normal conditions. *Physica A: Statistical Mechanics and its Applications*, 388(17):3600–3608, 2009. doi:10.1016/j.physa.2009.05.027.

[236] J. Chen, J. Ma, and S. M. Lo. Geometric constraint based pedestrian movement model on stairways. *Physica A*, 505:1212–1230, 2018. doi:10.1016/j.physa.2018.03.051.

[237] K. Lewin. *Field theory in social science: Selected theoretical papers.* Harper, New York, 1951.

[238] M. Chraibi, U. Kemloh, A. Schadschneider, and A. Seyfried. Force-based models of pedestrian dynamics. *Networks and Heterogeneous Media*, 6(3):425–442, 2011. doi:10.3934/nhm.2011.6.425.

[239] F. Dietrich, G. Köster, M. Seitz, and I. von Sivers. Bridging the gap: From cellular automata to differential equation models for pedestrian dynamics. *Journal of Computational Science*, 5(5):841–846, 2014. doi:10.1016/j.jocs.2014.06.005.

[240] T. Kretz. On oscillations in the Social Force Model. *Physica A: Statistical Mechanics and its Applications*, 438:272–285, 2015. doi:10.1016/j.physa.2015.07.002.

[241] P. G. Gipps and B. S. Marksjö. A micro-simulation model for pedestrian flows. *Mathematics and Computers in Simulation*, 27(2–3):95–105, 1985. doi:10.1016/0378-4754(85)90027-8.

[242] C. Burstedde, K. Klauck, A. Schadschneider, and J. Zittartz. Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A: Statistical Mechanics and its Applications*, 295:507–525, 2001. doi:10.1016/S0378-4371(01)00141-8.

[243] F. Dietrich and G. Köster. Gradient navigation model for pedestrian dynamics. *Physical Review E*, 89(6):062801, 2014. doi:10.1103/PhysRevE.89.062801.

[244] I. von Sivers and G. Köster. Dynamic Stride Length Adaptation According to Utility And Personal Space. *Transportation Research Part B: Methodological*, 74: 104–117, 2015. doi:10.1016/j.trb.2015.01.009.

[245] D. Hartmann. Adaptive pedestrian dynamics based on geodesics. *New Journal of Physics*, 12:043032, 2010. doi:10.1088/1367-2630/12/4/043032.

[246] M. J. Seitz and G. Köster. Natural discretization of pedestrian movement in continuous space. *Physical Review E*, 86(4):046108, 2012. doi:10.1103/PhysRevE.86.046108.

[247] C. M. Mayr and G. Köster. Social distancing with the Optimal Steps Model. *Collective Dynamics*, 6, 2021. doi:10.17815/CD.2021.116.

[248] C. A. S. Pouw, F. Toschi, F. van Schadewijk, and A. Corbetta. Monitoring physical distancing for crowd management: Real-time trajectory and group analysis. *PLoS ONE*, 15(10), 2020. doi:10.1371/journal.pone.0240963.

[249] I. Echeverría-Huarte, A. Garcimartín, R. C. Hidalgo, C. Martín-Gómez, and I. Zuriguel. Effect of physical distancing on the speed-density relation in pedestrian dynamics. *Journal of Statistical Mechanics: Theory and Experiment*, 2021 (4), 2021. doi:10.1088/1742-5468/abf1f0.

[250] U. Wilensky. NetLogo, 1999. URL `http://ccl.northwestern.edu/netlogo/`. Accessed on April 5, 2024.

[251] D. Adam. Special report: The simulations driving the world's response to COVID-19. *Nature*, 580(7803):316–318, 2020. doi:10.1038/d41586-020-01003-6.

[252] W. L. Oberkampf and C. J. Roy. *Verification and Validation in Scientific Computing*. Cambridge University Press, Cambridge, 2010.

[253] S. Funk, A. Camacho, A. J. Kucharski, R. Lowe, R. M. Eggo, and W. J. Edmunds. Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014–15. *PLOS Computational Biology*, 15(2):e1006785, 2019. doi:10.1371/journal.pcbi.1006785.

[254] M. Gödel, R. Fischer, and G. Köster. Applying Bayesian inversion with Markov Chain Monte Carlo to Pedestrian Dynamics. In *UNCECOMP 2019, 3rd ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering*, 2019. doi:10.7712/120219.6322.18561.

[255] T. J. Sullivan. *Introduction to Uncertainty Quantification*. Springer International Publishing, 1st edition, 2015. doi:10.1007/978-3-319-23395-6.

[256] M. A. Tumeo. *Stochastic and Statistical Methods in Hydrology and Environmental Engineering: Volume 2 Stochastic and Statistical Modelling with Groundwater and Surface Water Applications*, chapter The Meaning of Stochasticity, Randomness and Uncertainty in Environmental Modeling, pages 33–38. Springer Netherlands, 1994. doi:10.1007/978-94-011-1072-3_3.

[257] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity Analysis in Practice*. John Wiley & Sons, Ltd, 2004. doi:10.1002/0470870958.

[258] R. C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Computational Science and Engineering. Society for Industrial and Applied Mathematics, 2014. ISBN 978-1-611973-21-1.

[259] M. D. Morris. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2):161–174, 1991. doi:10.1080/00401706.1991.10484804.

[260] A. Saltelli and I. M. Sobol'. About the use of rank transformation in sensitivity analysis of model output. *Reliability Engineering & System Safety*, 50(3):225–239, 1995. doi:10.1016/0951-8320(95)00099-2.

[261] I. M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiment*, 1:407–414, 1993. doi:10.1016/S0378-4754(00)00270-6. English translation of the original paper [262].

[262] I. M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2:112–118, 1990.

[263] T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17, 1996. doi:10.1016/0951-8320(96)00002-6.

[264] I. M. Sobol'. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55:271–280, 2001. doi:10.1016/S0378-4754(00)00270-6.

[265] D. Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, Princeton, NJ, 2010. doi:10.2307/j.ctv7h0skv.

[266] I. M. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4): 86–112, 1967. doi:10.1016/0041-5553(67)90144-9.

[267] M. Gödel. *Systematic parameter analysis to reduce uncertainty in crowd simulations*. Dissertation, Technical University of Munich and Hochschule München, 2022.

[268] B. M. Adams, W. J. Bohnhoff, K. R. Dalbey, M. S. Ebeida, J. P. Eddy, M. S. Eldred, R. W. Hooper, P. D. Hough, K. T. Hu, J. D. Jakeman, M. Khalil, K. A. Maupin, J. A. Monschke, E. M. Ridgway, A. A. Rushdi, D. T. Seidl, J. A. Stephens, L. P. Swiler, A. Tran, and J. G. Winokur. Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.16 User's Manual. Online, 2022. URL `https://www.sandia.gov/app/uploads/sites/241/2023/03/Users-6.16.0.pdf`. Accessed on April 5, 2024.

[269] M. Baudin, A. Dutfoy, B. Iooss, and A.-L. Popelin. *OpenTURNS: An Industrial Software for Uncertainty Quantification in Simulation*, pages 1–38. Springer International Publishing, 2016. doi:10.1007/978-3-319-11259-6_64-1.

[270] J. Feinberg and H. P. Langtangen. Chaospy: An open source tool for designing methods of uncertainty quantification. *Journal of Computational Science*, Volume 11:46–57, 2015. doi:10.1016/j.jocs.2015.08.008.

[271] J. Herman and W. Usher. SALib: An open-source Python library for Sensitivity Analysis. *The Journal of Open Source Software*, 2(9), 2017. doi:10.21105/joss.00097.

[272] T. Iwanaga, W. Usher, and J. Herman. Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. *Socio-Environmental Systems Modelling*, 4:18155, 2022. doi:10.18174/sesmo.18155.

[273] S. da Veiga, F. Gamboa, B. Iooss, and C. Prieur. *Basics and Trends in Sensitivity Analysis.* Society for Industrial and Applied Mathematics, 2021. doi:10.1137/1.9781611976694.

[274] R. Ghanem, H. Owhadi, and D. Higdon, editors. *Handbook of uncertainty quantification.* Springer Cham, 2017. doi:10.1007/978-3-319-12385-1.

[275] J. H. Kwakkel. The Exploratory Modeling Workbench: An open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. *Environmental Modelling & Software*, 96:239–250, 2017. doi:10.1016/j.envsoft.2017.06.054.

[276] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, 2002. doi:10.1016/S0010-4655(02)00280-1.

[277] F. Campolongo, A. Saltelli, and J. Cariboni. From screening to quantitative sensitivity analysis. A unified approach. *Computer Physics Communications*, 182(4): 978–988, 2011. doi:10.1016/j.cpc.2010.12.039.

[278] A. B. Owen. On dropping the first Sobol' point. Online, 2020. doi:10.48550/ARXIV.2008.08051.

[279] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010. doi:10.1016/j.cpc.2009.09.018.

[280] M. J. W. Jansen. Analysis of variance designs for model output. *Computer Physics Communications*, 117(1):35–43, 1999. doi:10.1016/S0010-4655(98)00154-4.

[281] G. E. B. Archer, A. Saltelli, and I. M. Sobol'. Sensitivity measures, anova-like Techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58(2):99–120, 1997. doi:10.1080/00949659708811825.

[282] J. Schaffer. What Not to Multiply Without Necessity. *Australasian Journal of Philosophy*, 93(4):644–664, 2015. doi:10.1080/00048402.2014.992447.

[283] M. F. Lutfi. The physiological basis and clinical significance of lung volume measurements. *Multidiscip Respir Med.*, 12(1), 2017. doi:10.1186/s40248-017-0084-5.

[284] A. Chetta, A. Zanini, G. Pisi, M. Aiello, P. Tzani, M. Neri, and D. Olivieri. Reference values for the 6-min walk test in healthy subjects 20–50 years old. *Respiratory Medicine*, 100(9):1573–1578, 2006. doi:10.1016/j.rmed.2006.01.001.

[285] R. Larsen and T. Ziegenfuß. *Beatmung*. Springer Medizin, 4th edition, 2009. ISBN 9783540888123.

[286] A. F. Brouwer, M. H. Weir, M. C. Eisenberg, R. Meza, and J. N. S. Eisenberg. Dose-response relationships for environmentally mediated infectious disease transmission models. *PLOS Computational Biology*, 13(4):e1005481, 2017. doi:10.1371/journal.pcbi.1005481.

[287] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, Boston, MA, 1994.

[288] R. C. Martin. *Clean Code: A Handbook of Agile Software Craftsmanship*. Prentice Hall, 1st edition, 2008. ISBN 0132350882.

[289] J. C. Carver, N. Weber, K. Ram, S. Gesing, and D. S. Katz. A survey of the state of the practice for research software in the United States. *PeerJ Computer Science*, 8:e963, 2022. doi:10.7717/peerj-cs.963.

[290] M. Barker, N. P. Chue Hong, D. S. Katz, A.-L. Lamprecht, C. Martinez-Ortiz, F. Psomopoulos, J. Harrow, L. J. Castro, M. Gruenpeter, P. A. Martinez, and T. Honeyman. Introducing the FAIR Principles for research software. *Scientific Data*, 9(1), 2022. doi:10.1038/s41597-022-01710-x.

[291] H. Balzert. *Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements Engineering*. Springer Spektrum, Heidelberg, 2009. doi:10.1007/978-3-8274-2247-7.

[292] T. Richards. A review of software for crowd simulation. Online, 2020. URL `https://urban-analytics.github.io/dust/docs/ped_sim_review.pdf`. Leeds Institute for Data Science (LIDA), University of Leeds.

[293] I. Sommerville. *Software engineering*. Always learning. Pearson, 10th edition, 2016. ISBN 9781292096148.

[294] E. W. Dijkstra. The humble programmer. *Communications of the ACM*, 15(10): 859–866, 1972. doi:10.1145/355604.361591.

[295] G. Köster, M. Gödel, and S. Rahn. Modellierung und Simulation von Covid-Infektionsausbreitung in Menschenmengen in systemrelevanten Infrastrukturen : CovidSim-Projekt der Hochschule München : Schlussbericht zu Nr. 3.2 : Berichtszeitraum: 01.01.2021–31.12.2021. Technical report, Hochschule für angewandte Wissenschaften München, 2022. doi: 10.2314/KXP:1843060116.

[296] G. E. P. Box. *Robustness in the Strategy of Scientific Model Building*, pages 201–236. Academic Press, 1979. doi:10.1016/b978-0-12-438150-6.50018-2.

[297] G. E. P. Box. Science and Statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976. doi:10.1080/01621459.1976.10480949.

[298] K. Popper. *The Logic of Scientific Discovery (1934, 1959)*. Routledge Classics, London and New York, 2002.

[299] D. Fagundes. The Social Norms of Waiting in Line. *Law & Social Inquiry*, 42(04): 1179–1207, 2017. doi:10.1111/lsi.12256.

[300] A. Sieben, J. Schumann, and A. Seyfried. Collective phenomena in crowds — Where pedestrian dynamics need social psychology. *PLOS ONE*, 12(6):1–19, 2017. doi:10.1371/journal.pone.0177328.

[301] G. Köster and B. Zönnchen. A Queuing Model Based On Social Attitudes. In V. L. Knoop and W. Daamen, editors, *Traffic and Granular Flow '15*, pages 193–200, Nootdorp, the Netherlands, 2016. Springer International Publishing. doi:10.1007/978-3-319-33482-0_25.

[302] J. Schöttl, M. J. Seitz, and G. Köster. Investigating the Randomness of Passengers' Seating Behavior in Suburban Trains. *Entropy*, 21(6), 2019. doi:10.3390/e21060600.

[303] D. Lehmberg. *Operator-informed machine learning: Extracting geometry and dynamics from time series data*. Dissertation, Technical University of Munich and Hochschule München, 2022.

[304] T. Ishigami and T. Homma. An importance quantification technique in uncertainty analysis for computer models. In *[1990] Proceedings. First International Symposium on Uncertainty Modeling and Analysis*, pages 398–403, 1990. doi:10.1109/ISUMA.1990.151285.

[305] R. Saeedi, E. Ahmadi, M. S. Hassanvand, M. A. Mohasel, S. Yousefzadeh, and M. Safari. Implemented indoor airborne transmission mitigation strategies during COVID-19: a systematic review. *Journal of Environmental Health Science and Engineering*, 21(1):11–20, 2023. doi:10.1007/s40201-023-00847-0.

[306] M. Bicher, M. Wastian, D. Brunmeir, and N. Popper. Review on Monte Carlo Simulation Stopping Rules: How Many Samples Are Really Enough? *SNE Simulation Notes Europe*, 32(1):1–8, 2022. doi:10.11128/sne.32.on.10591.

[307] European Organization For Nuclear Research and OpenAIRE. Zenodo, 2013. doi:10.25495/7GXK-RD71.

A Infrastructure

My research relied on different resources and software packages. Most of them are open source or free. I owe thanks to the developers of these excellent tools. Table A.1 summarizes the software I used. Besides that, I would like to thank the authors of the template for this document.[1]

Table A.1: Overview of software that I primarily used for this dissertation.

Software	**Version**	**Link**
draw.io	15.4.0	`https://www.diagrams.net/`
Git	2.31.1	`https://www.git-scm.com/`
Inkscape	1.2.1	`https://inkscape.org/`
IntelliJ IDEA	2023.1.4	`https://www.jetbrains.com/idea/`
JabRef	5.11	`https://www.jabref.org/`
OpenJDK Java	17.0.2	`https://jdk.java.net/`
pdfTeX	3.141592653-2.6-1.40.24	`https://www.latex-project.org/`
PyCharm	2021.1.2	`https://www.jetbrains.com/pycharm/`
Python	3.8.2	`https://www.python.org/`
Texmaker	5.0.4	`https://www.xm1math.net/texmaker/`

[1] `https://github.com/TUM-LIS/tum-dissertation-latex`, commit 4329227, accessed on May 12, 2022

B Literature search: multidisciplinary interest in COVID-19

The document search in Section 2.2.2 was conducted on Scopus on January 2, 2024. The exact queries used for the document search were:

- For the number of documents by subject (see Fig. 2.10a):

```
TITLE-ABS-KEY ( {sars-cov-2} ) OR TITLE-ABS-KEY ( {covid-19} )
```

- For the number of documents by year (see Fig. 2.10b, search term **respiratory* and *disease**):

```
TITLE-ABS-KEY ( *simulat* OR *model* ) AND TITLE-ABS-KEY ( *
    respiratory* AND *disease* ) ) AND ( LIMIT-TO ( SUBJAREA , "
    COMP" ) OR LIMIT-TO ( SUBJAREA , "MATH" )
```

- For the number of documents by year (see Fig. 2.10b, search term *{sars-cov-2}*):

```
TITLE-ABS-KEY ( *simulat* OR *model* ) AND TITLE-ABS-KEY ( {sars-
    cov-2} ) ) AND ( LIMIT-TO ( SUBJAREA , "COMP" ) OR LIMIT-TO (
     SUBJAREA , "MATH" )
```

The search is case-insensitive and applies boolean operators. The asterisk represents a wildcard character. Curly braces force matching for exact phrases including punctuation. The data sets are listed in Appendix E (DS1). Acronyms used in the results are spelled out in Table B.1.

Table B.1: Acronyms used to summarize research areas in the document search

Acronym	Meaning
MEDI	Medicine
SOCI	Social Sciences
COMP	Computer Science
BIOC	Biochemistry, Genetics and Molecular Biology
ENGI	Engineering
IMMU	Immunology and Microbiology
ENVI	Environmental Science
BUSI	Business, Management and Accounting
PSYC	Psychology

C Literature search: modeling of infectious diseases

The systematic review of literature on mathematical modeling of infectious diseases in Section 3.2.2 is based on a document search on Scopus. The exact query used for the document search was:

```
TITLE-ABS-KEY ( model* OR simulat* ) AND TITLE-ABS-KEY ( pedestrian
    OR crowd OR "agent based" ) AND TITLE-ABS-KEY ( transmi* OR spread
    * ) AND TITLE-ABS-KEY ( disease OR virus OR viral OR pathogen OR
    infectio* ) AND TITLE-ABS-KEY ( air OR airborne OR aerosol OR
    droplet OR "*direct* transmi*" OR "contact transmission" ) ) AND
    PUBYEAR < 2024 AND ( LIMIT-TO ( PUBSTAGE , "final" ) ) AND ( LIMIT
    -TO ( DOCTYPE , "ch" ) OR LIMIT-TO ( DOCTYPE , "cp" ) OR LIMIT-TO
    ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) ) AND
     ( EXCLUDE ( EXACTKEYWORD , "Nonhuman" )
```

The purpose of this query was to determine the search range to be within agent-based models for transmission of respiratory diseases, for example, COVID-19. Table C.1 breaks down this specification. Note that strings such as COVID-19 and SARS-CoV-2 are too restrictive as conjunct values because they exclude literature published before the pandemic. As disjunct values, they are too broad because such keywords were frequently used after the outbreak. The search results are limited to articles, book chapters, or conference papers that have been published as of November 10, 2023, and are written in English. Furthermore, contributions labeled with nonhuman disease transmission are excluded. The raw data set is listed in Appendix E (DS1).

Table C.1: Search strings used in Scopus document search and their respective purpose.

No.	Specified search range	Search strings (row-wise conjunct)
S1	Field of modeling and simulation	model* *or* simulat*
S2	Agent-based models	pedestrian *or* crowd *or* "agent based"
S3	Infectious disease modeling	(transmi* *or* spread*) *and* (disease *or* virus *or* viral *or* pathogen *or* infectio*)
S4	Transmission paths associated with COVID-19 and similarly spreading diseases	air *or* airborne *or* aerosol *or* droplet *or* "*direct* transmi*" *or* "contact transmission"

The data set was further processed through manual selection during screening titles and abstracts followed by a full text review. Publications were excluded given that at least one of the following conditions was met:

- A description of or reference to the underlying model is missing.
- Human movement is not captured by an agent-based crowd dynamics model or trajectory data.
- The crowd model does not operate on a microscopic scale.
- The model does not target disease transmission.
- The model is designed for diseases that are transmitted other than via smaller or larger respiratory particles. Considering merely fomite transmission is not sufficient.
- The model does not refer to human-to-human transmission.
- The publication is accessible neither through open access or through my institutional log-in.

Table C.2: The table provides a non-exhaustive overview of disease transmission models coupled with microscopic crowd dynamics models. In this table, the persons are classified as susceptible (S), exposed (E), infectious (I), and removed or recovered (R). The transition between two states is indicated by an arrow. Further abbreviations: social force model (SFM), optimal steps model (OSM).

Ref.	Year	Disease transmission model	Crowd model	Scenario	Quantity of interest
		Proximity-based approaches			
[165]	2012	SIR-like model (see Eq. 3.1-3.3) $S \to I$: transmission probability $\kappa = 0.15$ within distance $d_e = 10\,\mathrm{m}$ to I $I \to R$: recovery at a certain rate $l = 3.2 \cdot 10^{-3}$	Real data	Fictitious disease spreading in an urban space; individuals commuting in central London, England	Proportion of S, I, and R over time
[11]	2014	$S \to I$: transmission within distance $d_e = 0\,\mathrm{m}$ to I	SFM	Fictitious disease spreading in an outdoor environment; crowd circling the Kaaba in Mecca, Saudi Arabia	Proportion of S and I over time
[166]	2014	$S \to E$: transmission is possible within distance $d_e = 1\,\mathrm{m}$	SFM	Fictitious disease spreading in a dense crowd passing through a corridor	Number of infections per time or distance walked
[167]	2017	Contact: established when mutual distance between two individuals is $< 0.46\,\mathrm{m}$ (physical contact, touching) or $< 0.76\,\mathrm{m}$ (potential transmission via larger respiratory particles) for $> 2.5\,\mathrm{s}$	SFM	Indoor environment; passengers boarding and deboarding an airplane following different boarding strategies	Number of contacts
[168]	2017	$S \to E$: Poisson distributed probability of infection within distance $d_e = 0.6\,\mathrm{m}$ to $2.1\,\mathrm{m}$, which may be attributed to larger and smaller respiratory particles, respectively	SFM	Indoor environment; transmission of Ebola virus in an airplane; one index patient at random position; limited validation by comparison to real data	Number of infected persons
[169]	2018	$S \to E$: see [167, 168], where $d_e = 1.2\,\mathrm{m}$ and $d_e = 2.1\,\mathrm{m}$ for transmission of Ebola virus via larger and SARS-CoV or influenza virus via smaller respiratory particles, respectively	SFM	Indoor environment; transmission of influenza virus (H1N1 strain), SARS-CoV, and Ebola virus among passengers of an airplane	Number of infected
[170]	2020	$S \to E$: see [167, 168]; smaller contact distances are associated with diseases such as Ebola, while larger distances are associated, for example, with influenza	SFM	Indoor environments; different strategies for passengers boarding or deboarding an airplane and queuing in front of a security check at the gate	Contact distances and times

[171]	2020	$S \to E$: see [167, 168], where $d_e = 0.9\,\mathrm{m}$ and $d_e = 2.1\,\mathrm{m}$ for transmission via larger and smaller respiratory particles, respectively	SFM	Indoor environment; transmission of influenza virus (H1N1 strain), SARS-CoV, and Ebola virus among passengers of an airplane and people queuing in front of a security check at the gate in different patterns	Number of infected
[172]	2020	Contact: probability of infection given that S and I are in contact, that is, the mutual distance is smaller than $d_e = 0.9\,\mathrm{m}$ to $2.1\,\mathrm{m}$ for exposure time $t_e > 4\,\mathrm{s}$	SFM	Transmission of pathogens such as influenza virus or SARS-CoV through larger and smaller respiratory particles between individuals queuing in different formations	Number of contacts and infections
[173], [174]	2020, 2022	$S \to I$: transmission within varying distances; increasing probability of infection the longer contact is maintained	OSM, network model	transmission of SARS-CoV-2 in various public spaces in Münster, Germany (local entities are connected through a network model)	Proportion of I over time; reproductive number
[175]	2020	$S \to E$: probability of infection (0.9) for S who stay within a distance of $d_e \leq 2.5\,\mathrm{m}$ for $t_e \geq 1\,\mathrm{min}$ or 2 min to I	Force-based	Indoor environment; spread of highly contagious pathogens, for example, measles virus among passengers moving through parts of William P. Hobby Airport in Houston, Texas, and during rides with shuttle buses between terminal and plane	Number of infected
[176]	2020	Contact: intensity measured as a negative exponential function of mutual distances between individuals accumulated over time	Velocity-based	Indoor environment; populations of different sizes and movement patterns shopping in a fictive supermarket	Contact intensity
[177]	2021	Exposure: accumulated exposure defined as $\int \exp(-\delta d(t))\,\mathrm{d}t$, where the spatial decay rate, δ, differs between talking ($16.29\,\mathrm{m}^{-1}$ for larger and $9.46\,\mathrm{m}^{-1}$ for smaller respiratory particles) and coughing ($7.64\,\mathrm{m}^{-1}$ for larger and $5.29\,\mathrm{m}^{-1}$ for smaller respiratory particles)	SFM	Indoor environment; transmission of SARS-CoV-2 within a crowd of 25 individuals exiting a room through a bottleneck of 1 m width; reference exposure resulting from a close contact ($d < 2\,\mathrm{m}$ for $t_e = 15\,\mathrm{min}$)	Maximum individual exposure
[178], [179]	2021, 2021	$I \to E$: probability of infection, P, as a linear function of the time S stay within distance of $d_e = 2\,\mathrm{m}$ to I; individual immunity levels of S and the efficiency of pathogen transmission and masks determine the slope of P	Scheduled activity (based on real data)	Indoor and outdoor environments; transmission of SARS-CoV-2 among individuals pursuing typical activities in a university building, on a cruise ship, and in an urban area;	Number of infected

[180]	2021	Exposure: exhaled and inhaled dose are described as functions of distance and headway orientations of I and S, respectively $S \to E$: probability of infection for a given inhaled dose is defined by a Wells-Riley equation	Real data	Outdoor and indoor environments; transmission of SARS-CoV-2 among stationary or dynamic pedestrians in public outdoor (streets, plaza, market, street cafés) and indoor (train or subway station) spaces in Lyon, France; densities between $0.038\,\mathrm{m}^{-2}$ to $0.46\,\mathrm{m}^{-2}$	Number of infected
[181]	2021	$S \to I$: individual probability of infection defined linearly as $\min(\beta t_e, 1)$ with transmission rate $\beta = 1.41 \cdot 10^{-9}$ and cumulative exposure time t_e during which S and I occupy the same zone (node)	Network model	Indoor environment; transmission of COVID-19 via larger respiratory particles in a supermarket	Exposure time and distance
[182]	2021	Exposure based on cumulative time S spends in proximity to ∗ I within distance $d_{e1} = 1\,\mathrm{m}$ ∗ cough disks around I with radius $d_{e2} = 2.5\,\mathrm{m}$, persistence time of 15 s (indoors) or 5 s (outdoors), and released by I every 15 s	SFM	Transmission of SARS-CoV-2 among moving pedestrians in generalized indoor and outdoor environments of $40\,\mathrm{m} \times 30\,\mathrm{m}$	Exposure time; number of persons with high-risk exposure time ($t_e > 20\,\mathrm{s}$)
[183]	2022	See [182]; adapted parameters for distance ($d_{e1} = 1.6\,\mathrm{m}$); cough disks persist for 5 s (outdoor conditions)	See [182]	Outdoor environment; Transmission of SARS-CoV-2 among pedestrians moving in a realistic railway station; different conditions for number of I, physical distancing, and mask wearing considered	Exposure time; number of persons with high-risk exposure time ($t_e > 15\,\mathrm{s}$)
[184]	2022	Exposure: S within distance $d_e = 1.7\,\mathrm{m}$ to coughing I experience certain exposure (derived from fluid dynamics simulation of a generic cough) Number of infected: linear function of the number of individuals with exposure above a threshold	SFM	Indoor environment; transmission of SARS-CoV-2 between individuals moving in a generic public space; only coughing considered	Individual exposure; number of infected
[185]	2022	Exposure: S accumulate exposure over time if their distance, d, to I is below $d_e = 3.5\,\mathrm{m}$ (without masks) or $d_e = 1.7\,\mathrm{m}$ (with masks); dose received during a time step $N_t \sim (1 - {}^{d}/_{d_e})^{2.5}$ $S \to E$: individual probability of infection for a given dose, N, is defined by $P = 1 - \exp(N)$	SFM	Indoor environments; transmission of SARS-CoV-2 among airplane passengers (boarding, deboarding, in-flight); re-enacted scenarios of in-flight spreading events; predictions under varying conditions, such as mask leakage and leaving seats vacant	Number of infected

[186]	2022	See [185]	See [185]	Indoor environments; transmission of SARS-CoV-2 among airplane passengers and visitors of a cinema	Number of infected
[187]	2022	$S \to E$: transmission is possible within distance $d_e = 1.5\,\mathrm{m}$	Random movement	Transmission of SARS-CoV-2 among individuals moving randomly in a generic area; time scale of 12 months	Number of infected
[188]	2022	$S \to E$: transmission within distance $d_e = 1\,\mathrm{m}$	Force-based	Indoor environment; transmission of SARS-CoV-2 among passengers traveling through Mianyang Airport, China; different percentages of people wearing masks are considered	Number of infected
[189]	2023	Exposure: cumulative time S spends at distance $d_e < 1.5\,\mathrm{m}$ to I	SFM	Indoor environment; populations of different sizes shopping in an average supermarket	Individual exposure time
[190]	2023	$S \to I$: probability of infection within certain distance to I	Random movement	Transmission of SARS-CoV-2 among individuals moving randomly in a generic area	
[191]	2023	Exposure: based on [180]; spatial mapping of the aerosol concentration improved by numerical calculation of particle transport; additionally, effect of modest winds or ambient airflows is taken into account	See [180]	See [180]	Number of infected
[192]	2023	SIR-like model; $S \to E$: transmission probability within distance $d_e = 2\,\mathrm{m}$ to I; $E \to I$: 4 to 6 days after exposure	SFM	Indoor and outdoor environments; transmission of SARS-CoV-2 among workers at a construction site; simulated time covers several days	Number of infected
[193]	2023	Exposure: based on [191]; (the preprint [193] was included in the targeted literature search; see [209] for the peer-reviewed version)	Various models (velocity-based, force-based, heuristics-based)	Viral transmission among pedestrians passing a corridor (bidirectional flow)	Number of infected per hour

Table continued on next page.

		Transmission medium-based approaches, e.g., exposure via aerosols or fomites			
[12]	2020	Exposure: modeled through exhalation and inhalation of respiratory particles; ∗ exhalation: continuous ($5\,\text{particles}\,\text{s}^{-1}$) and sporadic exhalation ($4 \cdot 10^4$ particles released at average frequency of $6\,\text{h}^{-1}$) exhaled particles immediately disperse uniformly within a volume of $1\,\text{m}^3$ ∗ aerosol concentration, $u\,(x,\,y,\,t)$: defined by 2D diffusion equation (Eq. 3.15), with diffusivity $D = 0.05\,\text{m}^2\,\text{s}^{-1}$ and sink term $-0.01\,\text{s}^{-1}u$; ∗ individual dose of inhaled aerosol particles: $N(t) = \dot{V}\int_0^t u\,(x,\,y,\,t)\,\text{d}t$ with individual inhalation rates $\dot{V} \sim \mathcal{U}\,(1.3,\,3.3)$ in $10^{-4}\,\text{m}^3\,\text{s}^{-1}$	Random movement (generic area); shortest path algorithm (supermarket)	Indoor environments; transmission of SARS-CoV-2 among stationary or slowly moving individuals in a generic indoor space and a fictive supermarket	Inhaled dose; number of individuals with critical doses $N > 100$
[15]	2021	Exposure: see [12], but different parameters for exhalation, air exchange, and inhalation rate	See [12]	Indoor environments; transmission of SARS-CoV-2 among individuals moving in a fictive supermarket or bar	Number of individuals with critical doses $N > 100$
[194] (preprint)	2021	$S \to E$: probability of infection modeled through an exponential dose-response relationship; the dose results from three transmission mechanisms with distinct properties for shedding, intake, and fate of respiratory particles; the viral load in the environment is mapped spatiotemporally by different layers for each transmission mechanism	Optimized utility [226]	Indoor environments; transmission of COVID-19 in restaurants	Cumulative viral contamination in the environment; relative contributions of transmission paths to individual exposure;
[195], [196], [197]	2021, 2021, 2022	Exposure: 3D, bi-directionally coupled model for airborne particle transport and crowd dynamics; particle transport is numerically calculated; focus on coughing or sneezing persons	SFM	Indoor environments; spread of airborne pathogens such as SARS-CoV-2 in scenarios where individuals walk along a corridor, pass through a long passage (train station), queue (fast food restaurant), and occupy a cabin (airplane / subway car); effects of ventilation and air cleaning through filtering (and ultra-violet radiation) are taken into account	Individual exposure; (number of infected [197])

[19]	2022	Exposure: based on semi-analytic description of airborne particle transport; background airflow patterns are defined manually	SFM	Indoor environment; transmission of SARS-CoV-2 between individuals leaving a building with multiple rooms; simulation time covers a few seconds	Individual exposure (times)
[17]	2022	Exposure: modeled through exhalation and inhalation of respiratory particles; * exhalation: I exhale spherical aerosol clouds at frequency $0.25\,\mathrm{s}^{-1}$ (breathing) with initial radius of $d_e = 1.5\,\mathrm{m}$ and pathogen load of 10^4 particles * aerosol concentration, u: concentration is instantaneously and homogeneously distributed within the volume of the cloud; clouds can overlap and increase when a person walks through the cloud; pathogen load decreases exponentially * inhalation: S inhale a fraction of the concentration at their position ($5 \cdot 10^{-4} u$) Exposure exceeding a threshold (obtained for a benchmark scenario of a close contact) is considered high-risk exposure	OSM	Indoor environment; airborne transmission of SARS-CoV-2 between individuals queuing in front of a service unit under different physical distancing conditions and reenacted super-spreading events occurring in a restaurant and during a choir rehearsal with one infectious person	Individual exposure, number of persons with high-risk exposure
[198]	2023	Exposure: modeled through exhalation and inhalation of respiratory particles; * exhalation: I contaminate their surrounding when speaking, coughing, or sneezing; these respiratory events have different properties (frequency of occurrence, shape of the contaminated area) * pathogen concentration: instantaneously and homogeneously distributed within $0.5\,\mathrm{m} \times 0.5\,\mathrm{m}$ grid cells; remains constant in time and space * inhalation: S inhale a fraction of the pathogen load at their position $S \to E$: individuals become infected when their exposure exceeds a fixed threshold	Random movement	Transmission of SARS-CoV-2 among pedestrians moving in a generic area of $30\,\mathrm{m} \times 30\,\mathrm{m}$; different types of masks are considered; the simulated time is limited to 10 min	Number of infected

[199]	2023	Exposure: modeled through exhalation and inhalation of respiratory particles; ∗ exhalation: I release different pathogen concentrations w when breathing, coughing, or sneezing; initial pathogen concentration at distance d to the exhaling individual's position is defined by $u_0(d) = {}^{w}/_{4}\,(1 + \cos(\pi {}^{d}/_{d_e}))$ for $d \leq d_e = 1\,\mathrm{m}$, otherwise $u_0 = 0$ ∗ aerosol concentration, $u(x, y, t)$: defined by 2D diffusion equation (Eq. 3.15), with diffusivity $D = 0.05\,\mathrm{m^2\,s^{-1}}$ and source / sink term $R_s = 0.585 u_0 - 9.25 \cdot 10^{-5}\,\mathrm{s^{-1}} u$ ∗ inhaled concentration, c: ${}^{\mathrm{d}c}/_{\mathrm{d}t} = \dot{V} u - Kc$, with intake rate $\dot{V} = 1.3 \cdot 10^{-4}\,\mathrm{m^3\,s^{-1}}$ and clearance rate $K = 1.9 \cdot 10^{-6}\,\mathrm{s^{-1}}$ $S \to E$: individual probability of infection is defined by $P = 1 - \exp\left(0.0069\,\mathrm{m^3\ copies^{-1}} c\right)$	SFM	Transmission of SARS-CoV-2 among 100 individuals moving randomly in an area of $50\,\mathrm{m} \times 50\,\mathrm{m}$; different parameter settings for interpersonal distance, desired walking speed, and age groups attempt to mimic typical Moroccan crowds in a shopping center, residential area, school, public space, and a workplace	Inhaled dose (resolved by age group)

Table continued on next page.

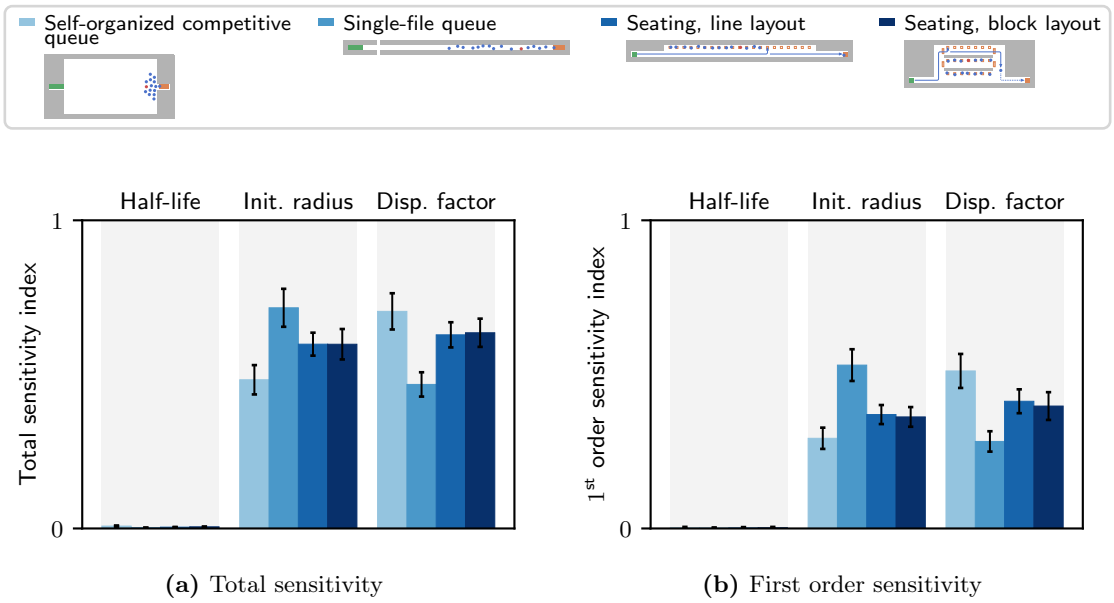
		Proximity- and transmission medium-based approaches			
[200]	2020	$S \to I$ based on ∗ proximity: probability following Bernoulli distribution within distance of 1 m ∗ medium: transmission via contaminated surfaces	SFM	Disease spreading in a closed population of 250 persons moving randomly in an area of 250 m × 250 m for 90 days	Number of infected (with different severity)
[13]	2021	Exposure based on ∗ proximity: exposure time of S increases when S and I occupy adjacent grid cells (cell size: 1 m × 1 m) ∗ medium: exhalation and inhalation of respiratory particles; continuous ($5\,\text{particles}\,\text{s}^{-1}$) and sporadic exhalation ($4 \cdot 10^4$ particles released at average frequency of $6\,\text{h}^{-1}$) yields immediately and uniformly distributed concentration within a volume of $1\,\text{m}^3$ (within a cell); particle concentration, u, decreases ($-u0.01\,\text{s}^{-1}$) and spreads spatially (not fully stated) over time; S continuously inhale a fraction of the pathogen load of a cell	Data-driven / cellular automaton	Indoor environment; transmission of SARS-CoV-2 among customers queuing at several service checkouts in a supermarket; variable availability of checkouts and compliance to physical distancing are considered	Total exposure time and number of inhaled particles of all persons
[201]	2021	$S \to E$ modeled for individuals moving on a grid of cells through ∗ proximity: transmission probability for S who are located in the same grid cell ∗ medium: I contaminate the cell they are located in with probability P_C; the cell remains contaminated for time t_R; S who enter a contaminated cell get infected with probability P_I	Random movement	Indoor or semi-open environments; transmission of pathogens such as SARS-CoV-2 between individuals moving randomly in an generic environment (square of grid cells); varying infection probabilities are considered for different age groups and infection states	Number of infected
[202]	2021	Exposure based on ∗ proximity: transmission probability $\kappa = 0$ to 0.1 within distance $d_e = 1\,\text{m}$ to I ∗ medium: I contaminate their immediate surrounding with probability p_1; S get exposed in contaminated areas with probability $p_2 = p_1 =$ 0 to 0.01; contamination level remains constant	SFM	Transmission of SARS-CoV-2 between 100 or 120 individuals moving randomly in an area of 30 m × 30 m and complying to different levels of physical distancing	Average exposure originating from distance- and location-based transmission

[18]	2022	Exposure based on * proximity: transmission within distance * medium (aerosols): I constantly emit pathogen; pathogens distribute instantaneously and homogeneously within a room; pathogen load decays over time; S inhale and, thus, remove a fraction of the pathogen load from the environment * medium (surfaces): similar concept as for transmission via aerosols but different spatial extent of contaminated areas and other parameter values	SFM	Indoor environments; transmission of SARS-CoV-2 among people carrying out everyday tasks in different parts of the University of Guadalajara, Mexico; interventions, such as face covering, physical distancing, and ventilation, are taken into account	Total exposure of all persons per transmission path
[14]	2022	Exposure modeled through exhalation and inhalation of respiratory particles based on * proximity: transmission via particles of different size ranges; mutual distances, head orientation of S and I, angle of the exhaled jet, and other factors are taken into account * medium: long-range transmission of fine aerosols; instantaneously and homogeneously distributed particle load within a room; air exchange is taken into account * $S \to E$: probability of infection for the normalized sum of all exposure levels, N, is defined by $P = 1 - \exp(-N)$	SFM	Indoor environment; transmission of SARS-CoV-2 among patients in a Chinese standard fever clinic	Individual exposure and infection probability
[203]	2023	$S \to E$ or I: probability of infection calculated based on cumulative exposure due to * proximity: exposure within distance $d_e = 1\,\mathrm{m}$ * medium (aerosols and surfaces): exposure within contaminated areas; areas for airborne transmission diffuse over time	SFM	Indoor environment; transmission of SARS-CoV-2 via respiratory particles and fomites among individuals navigating in an airport; different conditions, such as physical distancing and other movement patterns, are considered	Number of infected

Table continued on next page.

Other approaches						
[204]	2020	Exposure: accumulated and weighted exposure time evaluated based on distance, mutual orientation of individuals within certain reach, occupancy of the same room or building	–	–		Individual or averaged exposure time
[205], [206], [207], [208]	2020, 2020, 2020, 2021	Exposure in an aircraft cabin based on the * number of exposure contacts due to persons passing another person seated in the same row; * exposure time to possibly contaminated surfaces; * exposure time of a seated person to another person moving down the aisle and passing the row of the seated person; all individuals are potentially infectious	Cellular automaton	Indoor environment; focus on fomite and short-range transmission; evaluation of potential transmission risks during the boarding process of an airplane; different boarding strategies, locomotion behaviors, and COVID-19 mitigation strategies are considered		Total number of individuals exposed or total exposure time

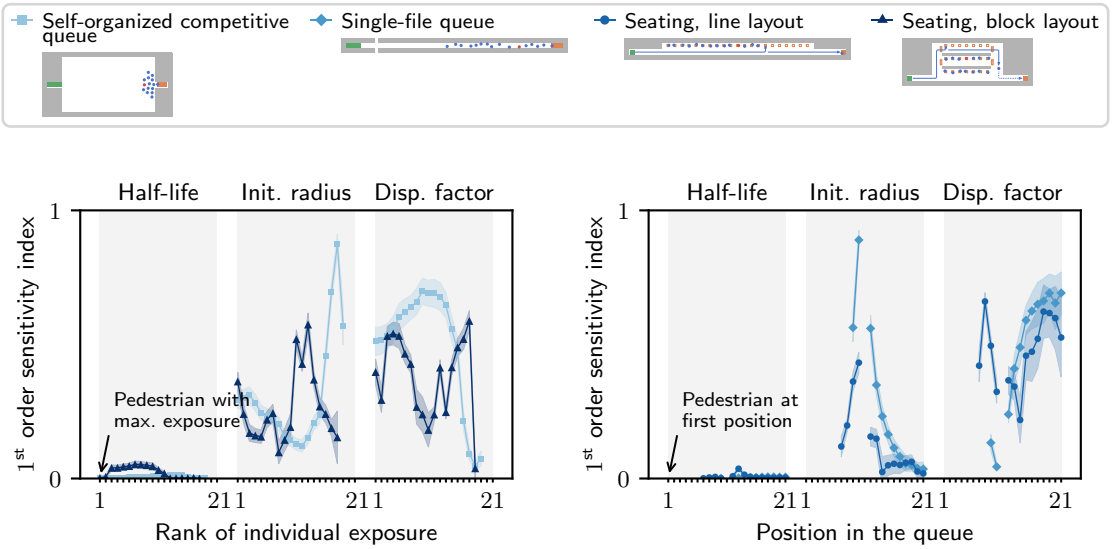
D Uncertainty quantification



(a) Total sensitivity

(b) First order sensitivity

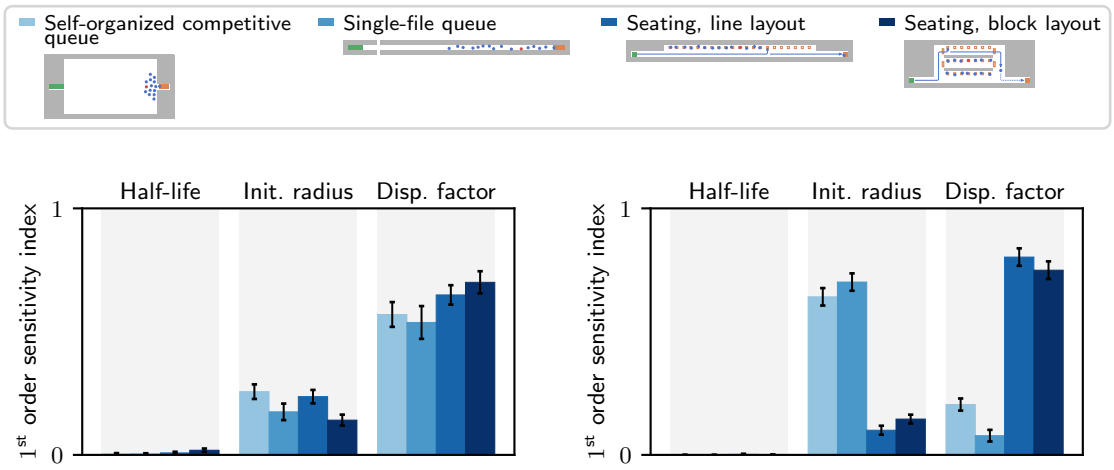
Figure D.1: The figures show the sensitivity indices of the maximum exposure occurring in a scenario with respect to the three uncertain parameters *half-life*, *initial radius*, and *dispersion factor*. The bootstrap confidence intervals (error bars) are computed from 100 re-samples at a 95 % confidence level.



(a) Quantity of interest: individual exposure ranked in descending order

(b) Quantity of interest: individual exposure sorted by queuing position in ascending order

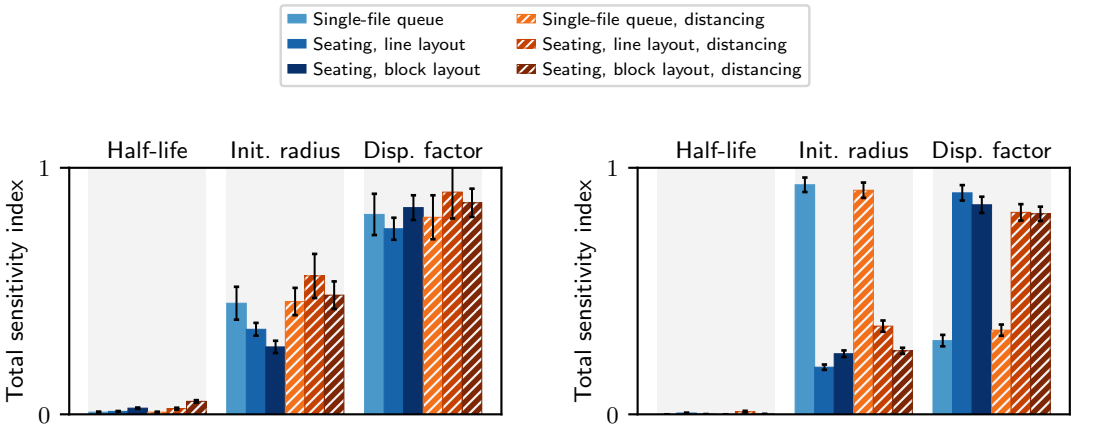
Figure D.2: The figures show the first order sensitivity indices of the individual exposure levels with respect to the three uncertain parameters *half-life*, *initial radius*, and *dispersion factor*. The sensitivity index of the infectious person is undefined because the respective exposure is always zero. It corresponds to rank 21 in (a) and to position 10 in (b). The bootstrap confidence intervals (shaded areas) are computed from 100 re-samples at a 95 % confidence level.



(a) Quantity of interest: average degree of exposure among the exposed persons

(b) Quantity of interest: number of exposed persons

Figure D.3: The figures show the first order sensitivity indices of the quantity of interest with respect to the three uncertain parameters *half-life*, *initial radius*, and *dispersion factor*. The bootstrap confidence intervals (error bars) are computed from 100 re-samples at a 95 % confidence level.



(a) Quantity of interest: average degree of exposure among the exposed persons

(b) Quantity of interest: number of exposed persons

Figure D.4: The figures compare the total sensitivity indices for the scenarios without physical distancing (shades of blue) and with physical distancing (shades of orange). The bootstrap confidence intervals (error bars) are computed from 100 re-samples at a 95 % confidence level. Note: own figures adapted from [213, p. 13].

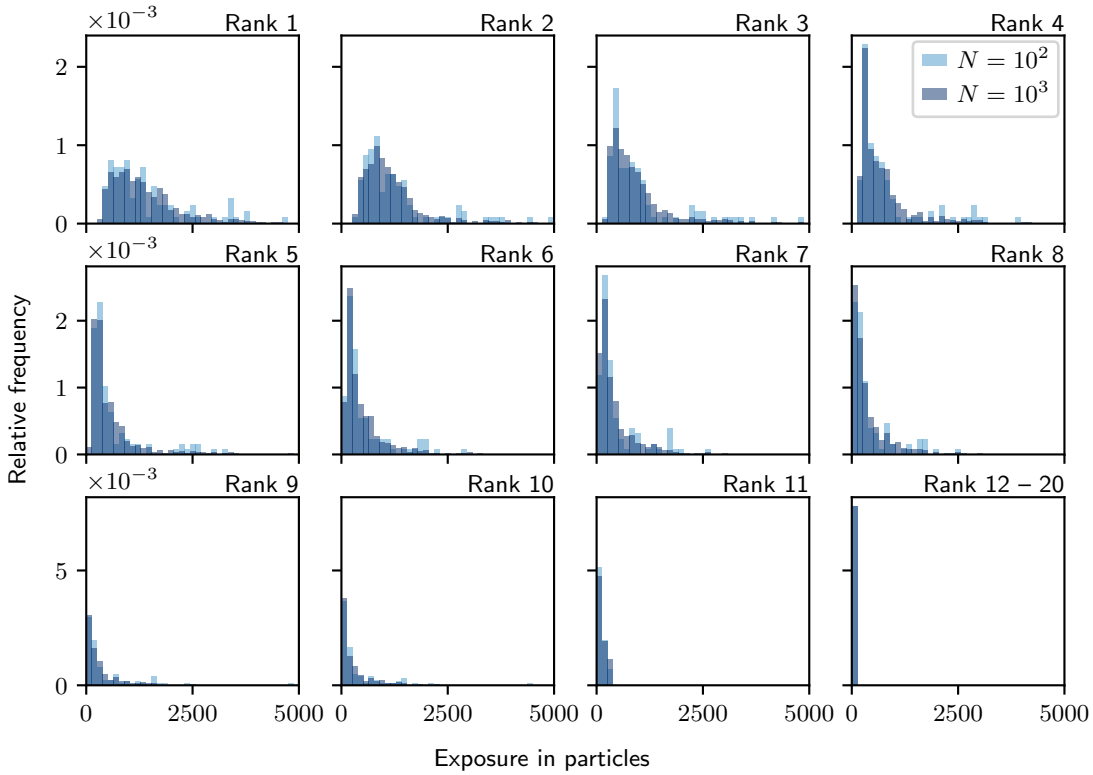


Figure D.5: The figure compares the result of Monte Carlo simulations with sample sizes of $N = 10^2$ and $N = 10^3$. The empirical distributions correspond to the ranks of individual exposure levels in the single-file queue obtained for a single repetition of the Monte Carlo experiment. The distributions for the ranks 12 to 20 are all the same because the respective exposure in the scenario is zero.

E Supplementary material

The supplementary material comprises data sets used in this dissertation. The files are hosted and archived in repositories managed by the European Organization For Nuclear Research and OpenAIRE [307].

DS1 Results of the literature search conducted in Section 2.2.2 / Appendix B:
`https://doi.org/10.5281/zenodo.10850095`

DS2 Results of the literature search conducted in Section 3.2.2 / Appendix C:
`https://doi.org/10.5281/zenodo.10850219`

DS3 Simulation data required for the comparison of crowd models in Section 3.2.3:
`https://doi.org/10.5281/zenodo.10845930`

DS4 Simulation data of the validation experiments in Section 7.1:
`https://doi.org/10.5281/zenodo.10846045`

DS5 Simulation data of the simulations analyzed in Section 7.2 and in my parameter study [213]:
`https://doi.org/10.5281/zenodo.10846072`

DS6 Simulation data of the uncertainty quantification experiments described in Chapter 8 and in my parameter study [213]:
`https://doi.org/10.5281/zenodo.10854747`