



## **Spatial heterogeneity affects predictions from early-curve fitting of pandemic outbreaks**

a case study using population data from Denmark

**Heltberg, Mathias L.; Michelsen, Christian; Martiny, Emil S.; Christensen, Lasse Engbo; Jensen, Mogens H.; Halasa, Tariq; Petersen, Troels C.**

*Published in:*  
Royal Society Open Science

*Link to article, DOI:*  
[10.1098/rsos.220018](https://doi.org/10.1098/rsos.220018)

*Publication date:*  
2022

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Heltberg, M. L., Michelsen, C., Martiny, E. S., Christensen, L. E., Jensen, M. H., Halasa, T., & Petersen, T. C. (2022). Spatial heterogeneity affects predictions from early-curve fitting of pandemic outbreaks: a case study using population data from Denmark. *Royal Society Open Science*, 9(9), Article 220018. <https://doi.org/10.1098/rsos.220018>

---

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Research



**Cite this article:** Heltberg ML, Michelsen C, Martiny ES, Christensen LE, Jensen MH, Halasa T, Petersen TC. 2022 Spatial heterogeneity affects predictions from early-curve fitting of pandemic outbreaks: a case study using population data from Denmark. *R. Soc. Open Sci.* **9**: 220018. <https://doi.org/10.1098/rsos.220018>

Received: 18 January 2022

Accepted: 16 August 2022

**Subject Category:**

Mathematics

**Subject Areas:**

mathematical modelling/biophysics/  
computational biology

**Keywords:**

pandemics, agent-based modelling,  
spatial heterogeneity, fitting, COVID-19

**Author for correspondence:**

Mathias L. Heltberg

e-mail: [heltberg@nbi.ku.dk](mailto:heltberg@nbi.ku.dk)

# Spatial heterogeneity affects predictions from early-curve fitting of pandemic outbreaks: a case study using population data from Denmark

Mathias L. Heltberg<sup>1,2,3,†</sup>, Christian Michelsen<sup>1,†</sup>,  
Emil S. Martiny<sup>1,†</sup>, Lasse Engbo Christensen<sup>4</sup>, Mogens  
H. Jensen<sup>1</sup>, Tariq Halasa<sup>5</sup> and Troels C. Petersen<sup>1</sup>

<sup>1</sup>Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, Copenhagen E 2100, Denmark

<sup>2</sup>Laboratoire de Physique, Ecole Normale Supérieure, Rue Lhomond 15, Paris 07505, France

<sup>3</sup>Infektionsberedskab, Statens Serum Institute, Artillerivej, Copenhagen S 2300, Denmark

<sup>4</sup>DTU Compute, Section for Dynamical Systems, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Anker Engélunds Vej 101A, Kongens Lyngby 2800, Denmark

<sup>5</sup>Animal Welfare and Disease Control, University of Copenhagen, Grønnegårdsvej 8, Frederiksberg C 1870, Denmark

MLH, 0000-0002-9699-4075; LEC, 0000-0001-5019-1931

The modelling of pandemics has become a critical aspect in modern society. Even though artificial intelligence can help the forecast, the implementation of ordinary differential equations which estimate the time development in the number of susceptible, (exposed), infected and recovered (SIR/SEIR) individuals is still important in order to understand the stage of the pandemic. These models are based on simplified assumptions which constitute approximations, but to what extent this are erroneous is not understood since many factors can affect the development. In this paper, we introduce an agent-based model including spatial clustering and heterogeneities in connectivity and infection strength. Based on Danish population data, we estimate how this impacts the early prediction of a pandemic and compare this to the long-term development. Our results show that early phase SEIR model predictions overestimate the peak number of infected and the equilibrium level by at least a factor of two. These results are robust to variations of parameters influencing connection distances and independent of the distribution of infection rates.

<sup>†</sup>These authors contributed equally.

# 1. Introduction

Over the past years, the pathogen now known as SARS-CoV-2 has spread dramatically, risen in several waves, paralyzing societies, resulting in a large number of deaths and severe economic damage worldwide [1,2]. Mathematical models have estimated the reproduction number and guided the authorities in an attempt to minimize the damage caused by the virus [3–6]. Even though modern algorithms using machine learning have helped the process [7,8], the majority of models used to predict the size of the pandemic (or a rising wave of the disease) have been variants of the SIR/SEIR model. The SIR model was originally proposed in 1927, in the seminal work of Kermack and McKendrick, who successfully described the evolution of a pandemic, using a mean field approximation where all individuals are described as one population [9]. In the investigations of the SARS-CoV-2 pandemic, the mathematical models have varied in complexity including simple deterministic compartmental models [6,10], meta-population compartmental models [11–13], individual based models without including spatial specifications [4,14,15] and spatio-temporal agent-based models [16].

One aspect in the modelling is the ability to predict the infection peak height and the number of individuals who will be infected based on the early rise in the number of infected (before governmental interference). Earlier work has pointed out the importance of including heterogeneity when modelling the spread of infectious disease such as contact patterns between individuals [17], population mixing assumptions [18], heterogeneities caused by super-spreaders [15], and the spatial dependency of COVID-19 [19,20]. These mathematical models have not combined heterogeneous elements nor quantified how much the early SIR/SEIR predictions might be biased.

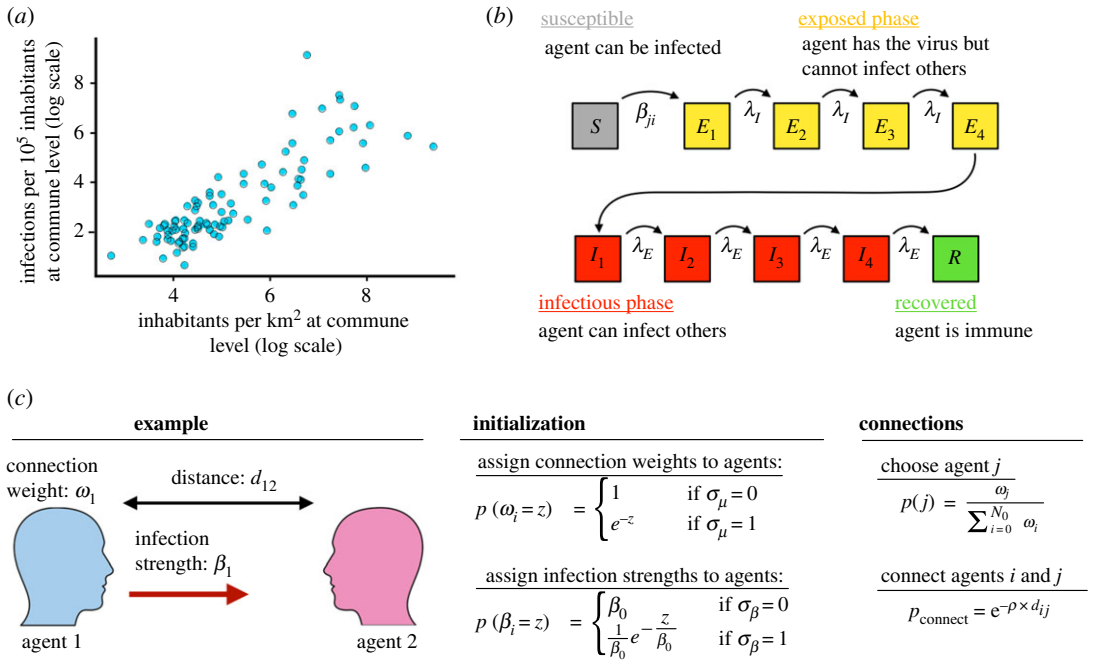
In this paper, we include geographical distributions based on an entire population, using population data of Denmark. When the SIR model was originally formulated, 95 years ago, data was not available to investigate the effects of geographical and demographic differences among the population, which might be one of the reasons why fundamental properties for diseases, such as the basic reproduction number ( $R_0$ ), can vary significantly between different regions [21]. However, with modern collection of data, these geographical aspects might be accounted for. Our main goal of this work is therefore to investigate the importance of heterogeneities in a geographically distributed population on the spread of a pandemic. We find that the heterogeneity arising from spatial inhomogeneities causes an increase in the early stage of the pandemic, affecting the initial forecast and highlighting the importance of early intervention in order to minimize the effects of the pandemic.

## 1.1. Construction of the model

In order to investigate the effect of a geographically distributed population, we extracted the number of infected per commune (from the Danish Serum Institute [22]) and divided this number with the number of inhabitants in each commune to obtain the number of infected per individual in each commune. This number we then plotted against the number of inhabitants in that specific commune (extracted from statistics Denmark [23]). Doing so, we found a strong correlation between the population density and the number of infections per inhabitant as seen in figure 1*a*. This observation has been made for many other countries [24–29] and underlines the aspect of disease spreading that has been observed since ancient times; that densely populated regions often have larger pandemics than the rural areas. Note that in the very early stage of a pandemic, before the exponential growth rate is reached, micro outbreaks will guide its evolution and these events can likely take place in regions with low density [30].

## 1.2. Disease simulation

To simulate evolution of the disease, we assigned each individual (agent) to a state (predominantly initialized in state *S*) and assigned four states to the exposed phase and four states to the infectious phase, in order to achieve an Erlang distribution (which is related to the Gamma distribution) of time in each phase [31]. Once in the exposed phase, the infected agent has a rate to move into another state, where the rate is fixed based on experimental data in order to achieve a mean time in the exposed phase of approximately 4 days (table 1). Each agent in the Infectious phase can infect other agents that have a connection to this agent in the network. This definition of agents in discrete states is naturally a simplification of the real pandemic, and we stress that this mathematical model aims at describing the spread of the disease in a simple way that does not capture all aspects of the real disease. We do not believe that this impacts our main conclusions in any way, as we are aware that one should always be careful when making these kinds of simplifications. To investigate the effect of



**Figure 1.** (a) Population density ( $x$ -axis) and the number of infections per  $10^5$  inhabitants ( $y$ -axis) for each commune in Denmark. (b) Illustration of the modified susceptible-exposed-infected-removed (SEIR) model used. It consists of 10 consecutive states ( $S, E_{1-4}, I_{1-4}$  and  $R$ ), with transition rates governed by  $\beta, \lambda_E$  and  $\lambda_I$ , respectively. (c) Illustration of how the spatial network is generated and heterogeneities in individuals included.

infection heterogeneities, we assigned an infection strength to each connection in the network, so some agents were more infectious than others. In order to control the degree of this heterogeneity, we assigned a boolean parameter  $\sigma_\beta$ , that if switched on generated an exponential distribution in infection strengths, keeping the mean field reproduction number fixed. The reproduction number between the ABM and the SIR model is related through the parameter  $\tilde{\beta} = \beta(\mu/2N_0)$ . All transitions between states and infection of other individuals were done using the Gillespie algorithm [32]. This is schematized in figure 1b.

### 1.3. Network creation

In order to construct the underlying network, we created a set-up whereby two agents were chosen at random but based on their individual connectivity weight each iteration and connected with some probability based on their spatial position. To include the possibility of highly connected individuals independent of their spatial position, we assigned a boolean parameter  $\sigma_\mu$ , that, if switched on, generated an exponential distribution in weights for the individuals, keeping the mean field reproduction number fixed similar to the heterogeneity in infection strengths. To include the spatial position in the network, we introduced a parameter  $\rho$ , so the probability of connecting two chosen agents decayed exponentially with the distance between them:  $p_{\text{connect}} = e^{-\rho \times d_{ij}}$ . In order to allow some long-distance connections we introduced another parameter  $\epsilon \in [0, 1]$ , that determines the fraction of distance-independent contacts. To construct the network of spatially distributed contacts, we chose the parameters using data based on:

- The geographical location of people in Denmark (from Boligsiden [33])
- The average number of contacts per individual per day of 11 (from HOPE [34]). Given an average infectious period of 4 days, we approximate the average number of effective contacts to be  $\mu = 40$
- The average commuting distance  $\rho = 0.1 \text{ km}^{-1}$  and the fraction of long-distance commutes  $\epsilon_\rho = 4\%$  (from statistics Denmark [23])

This is schematized in figure 1c and further described in the Methods section. All 10 parameters in this model are defined and outlined in table 1. We note that in order to keep the parameters space low, this model does not include the effects of temporal changes such as seasonality and holidays. While all agents

**Table 1.** Overview of the 10 parameters applied in this study, their typical value, and the ranges we have considered. The first six parameters are standard SEIR parameters, whereas the last four parameters define the heterogeneity in the model. These four parameters do not affect the SEIR model.

variable	description	value	range	units
$N_0$	population size	$5.8 \times 10^6$	$10^5$ – $10^7$	—
$N_{\text{init}}$	number of individuals initially infected	100	$1$ – $10^4$	—
$\mu$	average number of network contacts	40	10–100	—
$\beta$	typical infection strength	0.01	0.001–0.1	$\text{d}^{-1}$
$\lambda_E$	rate to move through $\frac{1}{4}$ of latency period	1	0.5–4	$\text{d}^{-1}$
$\lambda_I$	rate to move through $\frac{1}{4}$ of infectious period	1	0.5–4	$\text{d}^{-1}$
$\sigma_\mu$	population clustering spread	0	0–1	—
$\sigma_\beta$	interaction strength spread	0	0–1	—
$\rho$	typical acceptance distance	0.1	0–0.5	$\text{km}^{-1}$
$\epsilon_p$	fraction of distance-independent contacts	0.04	0–1	—

have been assigned parameters to their infection network that are derived from statistics of Denmark for both employees and students, we have not divided each agent into specific occupations.

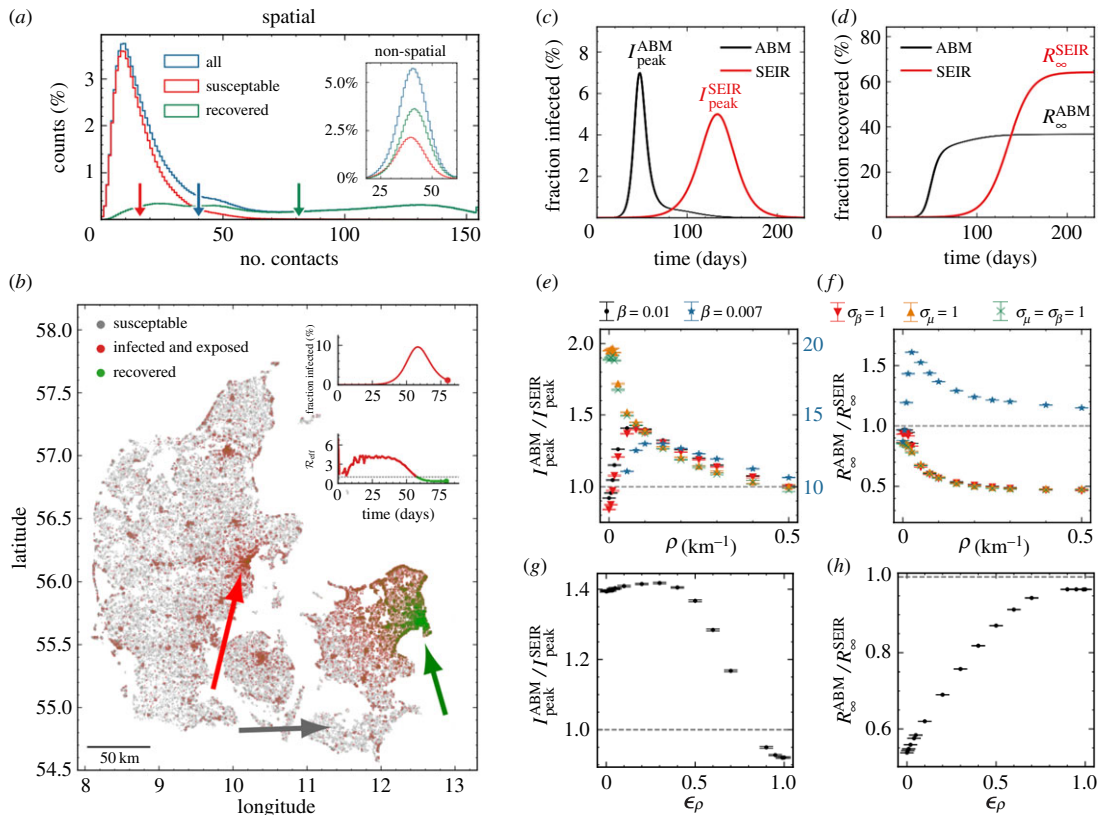
Before including heterogeneity, we compared the ABM to the corresponding SEIR model as a test, and found them to agree within 5% for all parameter configurations tested. Here, we also tested the effect of the number of individuals initially infected (see electronic supplementary material). This concludes that the SEIR and ABM model are calibrated to have the same reproduction number in the absence of heterogeneities. Next, we will introduce heterogeneities into the system, while keeping the sum of contacts and infection strengths constant, to study how this affects the evolution of the pandemic.

## 2. Results

### 2.1. Geographical distributions in a population and large variances in numbers of contacts leads to increased infection levels

Having introduced heterogeneity, the distributions of connections in this network were created automatically through the population clustering, see figure 2*a*. This naturally leads to individuals living in densely populated areas having higher number of connections. In an example simulation with 100 initially infected individuals,  $N_{\text{init}} = 100$ , we observed a spatial difference in areas affected by the disease (figure 2*b*), as expected. Note that we also show the effective reproduction number ( $\mathcal{R}_{\text{eff}}$ ) as a function of time in the lower part of the inserted panel. One region reached local endemic steady state (green arrow, figure 2*b*) while other regions of similar density were highly infected (red arrow, figure 2*b*) and yet other districts were almost unaffected (grey arrow, figure 2*b*). To quantify the effect of population clustering, we compared the ABM result to the reference SEIR model of similar parameters. Generally, we observed that the epidemic developed faster with a higher infection peak  $I_{\text{peak}}$ , but also subsided quicker, leading to a lower number of infected once reaching endemic steady state,  $R_\infty$  (figure 2*c,d*).

In order to explore how population clustering affects the epidemic, we chose a reference value of infection rates,  $\beta = 0.01$ , and an alternative value of  $\beta = 0.007$ . In the absence of spatial dependence ( $\rho = 0 \text{ km}^{-1}$ ), these correspond to initial reproduction numbers  $\mathcal{R}_0 \approx 1.7$  and 1.1, respectively. Here, we define the reproduction number as the average number of agents each infectious agent will infect in the first part of the disease. Increasing the spatial dependence (i.e. increasing  $\rho$ ) leads to a significant rise in the infection peak for the ABM,  $I_{\text{peak}}^{\text{ABM}}$ , compared to the (unaffected) SEIR model,  $I_{\text{peak}}^{\text{SEIR}}$  for both the reference value and the alternative lower value of  $\beta$  (black and blue points, figure 2*e*). We introduced heterogeneity in infection strengths ( $\sigma_\beta = 1$ , see figure 1*b*), thus making some individuals much more infectious than others (i.e. including *super shedders*). We found no significant impact from this effect (red points in figure 2*e*). Similarly, we introduced heterogeneity in connection weights ( $\sigma_\mu = 1$ , see figure 1*b*), thus making some individuals much more likely to form contacts than others



**Figure 2.** (a) Histograms showing the number of susceptible (red) and recovered (green) individuals at the end of an epidemic with  $\rho = 0.1 \text{ km}^{-1}$ . The distribution before the epidemic is shown in blue. The arrows show the mean of each distribution. The inset shows the same for  $\rho = 0 \text{ km}^{-1}$ . (b) Visualization of the spatial position of individuals during the infection and which state they are in. Green arrow: largest city in Denmark (Copenhagen): mostly recovered. Red arrow: Second largest city in Denmark (Aarhus): mostly infected. Grey arrow: low-population area: mostly susceptible (i.e. have not been infected). (c) Number of infected individuals as a function of time. Data shown for the spatially distributed network ( $\rho = 0.1 \text{ km}^{-1}$ ). Simulation was repeated 10 times. (d) Cumulative sum of individuals who have had the disease as a function of time (with  $\rho = 0.1 \text{ km}^{-1}$ ). (e) Relative difference in maximal number of infected,  $I_{\text{peak}}$ , between deterministic (SEIR) and ABM as a function of  $\rho$ , and shown for different parameters. Note the data for  $\beta = 0.007$  are shown in blue with a factor 10 scaling (right y-axis). (f) Relative difference in total number of infected at the end of the epidemic,  $R_{\infty}$ , between deterministic (SEIR) and ABM as a function of  $\rho$ . Colours similar to (e). (g) Same as (e), but as a function of  $\epsilon_{\rho}$ . (h) Same as (f), but as a function of  $\epsilon_{\rho}$ .

(i.e. including *super connectors*). This leads to a significant effect for  $\rho = 0 \text{ km}^{-1}$ , which converges towards the other curves for  $\rho > 0.1 \text{ km}^{-1}$  (orange (only super connectors) and green (super connectors and super shedders) points in figure 2e). The total number of individuals that have been in the infectious state, when there are not enough susceptible agents for the disease to keep infecting new individuals, is termed  $R_{\infty}$ , and this converged towards half of the SEIR model prediction as a function of  $\rho$  except for  $\beta = 0.007$  where the endemic steady state level is larger than the one obtained by the SEIR model (figure 2f). We note that in reality, individuals can lose immunity and therefore new waves can emerge. But for a completely susceptible population,  $R_{\infty}$  describes the fraction of the population that will get the disease during a specific wave. Fixing  $\rho = 0.1 \text{ km}^{-1}$  and increasing the fraction of distance-independent contacts,  $\epsilon_{\rho}$ , we found that  $I_{\text{peak}}^{\text{ABM}}$  is almost unaffected for  $\epsilon_{\rho} < 0.5$  (figure 2g), while  $R_{\infty}^{\text{ABM}}$  increases linearly towards the SEIR model  $R_{\infty}^{\text{SEIR}}$ , as expected (figure 2h).

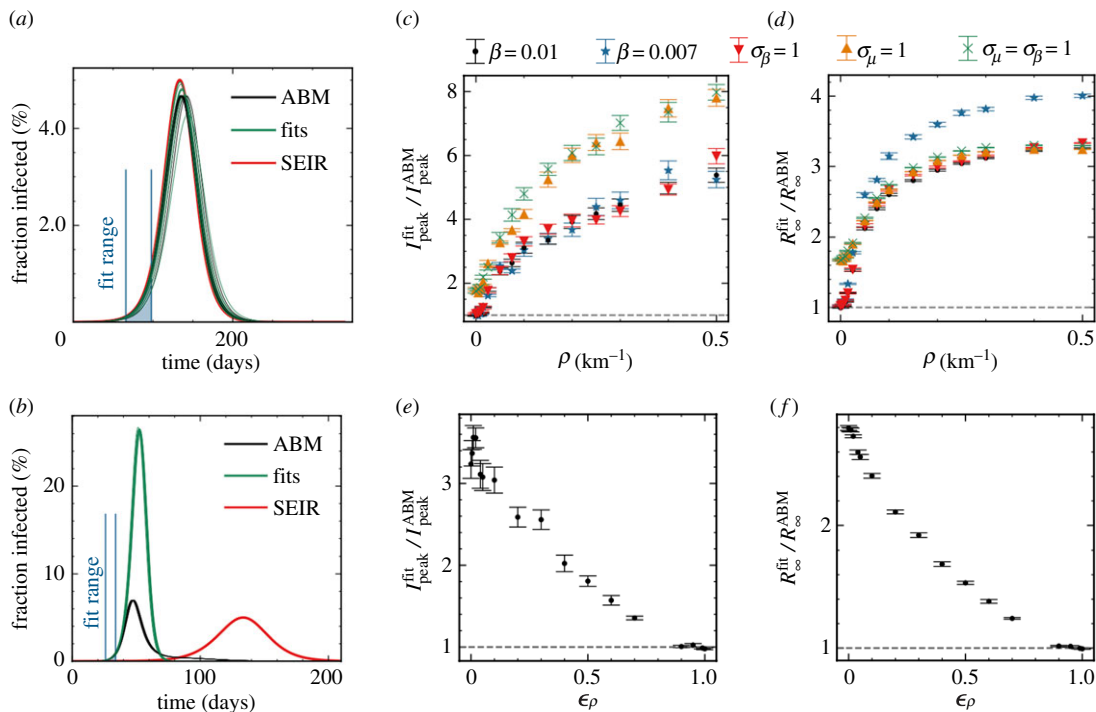
## 2.2. Fitting early infection curves leads to significant bias in estimating the size of the pandemic

Next, we consider how these heterogeneities bias the traditional SEIR model predictions, especially the predictions based on fits to the number of infected (i.e. the curve to be flattened) in the beginning of the epidemic (see Methods). Without spatial dependence, the predicted curves fitted the number of infected

individuals very well (figure 3a). Introducing spatial dependence ( $\rho = 0.1 \text{ km}^{-1}$ ) leads to a severe overestimation of the epidemic based on the number of early infection cases (figure 3b). This result can be interpreted by the fact that in societies where population density and thus individual contact number varies significantly, the early phase will be driven by people with many contacts (*super connectors*). This typically happens in cities where the population density is high. Increasing the spatial dependence  $\rho$ , we found that the SEIR model predictions overestimated the infection peak height  $I_{\text{peak}}$  and the total number of infected  $R_{\infty}$  significantly even for very small spatial heterogeneities (figure 3c, d). We observed this general trend for all tested combinations of parameters and heterogeneities. In particular, we found that if long-distance connections  $\epsilon_p$  are below 10%, the bias in the estimated infection peak height,  $I_{\text{peak}}$ , was constant within statistical uncertainty (figure 3e). For the total number of infected,  $R_{\infty}$ , we observed an almost linear regression to the SEIR model as  $\epsilon_p$  approaches one. However, even when  $\epsilon_p = 0.25$ , the prediction bias was still a factor of two (figure 3f). We concluded from these curves a general trend; if one fits an SEIR model to infection numbers during the beginning of an epidemic, and use these estimates to predict the characteristics of the epidemic at a national level, one overestimates the number of infected by at least a factor of two.

### 3. Discussion

In summary, this work outlines that the degree of population clustering in Denmark creates a discrepancy between the early predictions made by the SEIR models and the underlying agent-based interactions. It results in a significant overestimation of the impact of the disease, both in terms of maximal number of simultaneously infected (by a factor of 3) and the endemic steady state level (by a factor of 2.5). Such discrepancies have been observed for earlier pandemics, for instance, the 1918 Spanish flu, where the predicted number of individuals that would get the disease within a season turned out to be higher than the actual outcome [35]. The present results can be an important element in explaining these mismatches, even though other elements, such as for instance social distancing and the population behaviour, play a vital part. When facing a rising pandemic, societies are faced with the task of laying out strategies to minimize the consequences, including the importance of *flattening the curve*. While this is truly crucial to avoid overpopulated hospitals, the understanding of the pandemic should be taken seriously enough that we might specify to a higher degree of certainty which curve to be flattened. Our results highlight an important element in the prediction of infection levels and quantify the effect of density heterogeneities. We are aware that these results are not directly applicable to the pandemic of SARS-CoV-2 as a whole, since numerous mutations have increased the infection rates compared to the early estimates and created a strong heterogeneity in the infection worldwide. Furthermore, the actual evolution of the pandemic was highly affected by the different governmental interventions, that are not included in this work. However, this study emphasizes the abnormally large reproduction rates in the beginning of a pandemic, due to individuals with more connections than the rest of the population and attempts to quantify this bias, when countries should estimate the severity of a disease based on the data collected in the early phase. This also underlines the benefits by making lockdowns early in the pandemic, when a population is highly susceptible (for instance to a new mutation) and therefore can be driven by *super connectors*. Since people living in city-clusters are more likely to have many contacts, or infection events, they are on average more likely to be affected in the early stage of the pandemic (if they do not implement social distancing). By removing contacts from these individuals, through some level of interaction in order to reduce the number of social contacts, one can avoid the worst peak while affecting the lowest number of people. While our work describes some fundamental aspects of the disease spreading, this model does not consider asymptomatic individuals, which has been an important aspect of the SARS-CoV-2 pandemic [36,37]. Effectively, asymptomatic individuals would correspond to a very heterogenous distribution of time the agents spend in the infectious state. While agents with symptoms would predominantly isolate themselves and thereby significantly reduce their ability to infect other agents, asymptomatic agents would have a long time in the infectious state, thereby infecting more individuals. In this work, we have not considered the observation that individuals lose their immunity to SARS-CoV-2 which was first studied in the Brazilian city of Manaus. For this model, the temporal decline of immunity would lead to more pandemic ‘waves’, but for a fixed disease transmissibility this would not alter the maximal height of the peak number of infected, since this occurred for all the initially susceptible population. Finally, we note that this work does not include a vast range of divisions for the population, including age, socio-economic status etc. We have not included this directly, since we wanted to estimate as cleanly as possible how the heterogeneity in the



**Figure 3.** (a) Number of infected individuals for the ABM in black, the SEIR model in red and the SEIR fits to the ABM data in green. Blue lines show the interval where parameters are fitted (also shown below the curves). Here,  $\rho = 0 \text{ km}^{-1}$ . (b) Same as (a) but with population clustering ( $\rho = 0.1 \text{ km}^{-1}$ ). (c) Relative difference in maximal number of infected,  $I_{\text{peak}}^{\text{fit}} / I_{\text{peak}}^{\text{ABM}}$ , between the fit and the ABM for different values of  $\rho$ . Simulations repeated 10 times for each data-point. (d) Relative difference in total number of infected at the end of the epidemic,  $R_\infty^{\text{fit}} / R_\infty^{\text{ABM}}$ , between the fit and the ABM for different values of  $\rho$ . (e) Same as (c), but as a function of  $\epsilon_\rho$ . (f) Same as (d), but as a function of  $\epsilon_\rho$ .

contact pattern, arising from a geographically distributed population, could affect the evolution of a disease. We are aware that for instance the distribution of age has an enormous impact on the health risk and that this risk is vital in the prediction of hospitalizations in modern society. However, our aim was to understand the bias in the prediction of a disease, based on the data that comes during the early periods of a disease, independently of the mortality of this disease. Mathematical predictions of disease progression have been heavily criticized [38,39] and it is important to improve the theoretical foundations of the mathematical descriptions, in order to increase the confidence in the predictions. Our work highlights the importance of estimating the spatial clustering and connectivity skewness in the population in order to correct the predictions based on SEIR models, by quantifying their biases from not including spatial clustering. We hope that this work could serve as an input to the modelling and prediction of future pandemics and the importance of avoiding super-spreaders in high-density areas.

## 3.1. Methods

### 3.1.1. Construction of spatial network

We initialized  $N_0$  agents on a network generating a total of  $\mu \times N_0$  links between two agents, with an assigned interaction strength  $\beta_{ij}$  for each link. The average contact number,  $\mu$ , was fixed to 20, based on results from the Danish HOPE project, gathering data on population behaviour since April 2020 [34]. In order to include a realistic, geographical distribution of the population, we randomly selected agent locations from a two-dimensional kernel density estimate we had generated based on housing sales in Denmark 2007–2019 (data given with permission from Boligsiden, [33]). We note that in this distribution, we do not take specific geographical elements such as roads or environment into account (which has been previously studied for other diseases [40]) as we assume that this effect is small in a country like Denmark, where all parts are connected and natural obstacles such as mountains and rivers are not present. To connect the agents, we used a hit and miss method, where two random agents are first suggested and then connected with probability,  $p(d) = e^{-\rho d_{ij}}$ . Here,  $d_{ij}$  is the distance between agents and



$\rho$  is a parameter with units of inverse distance. We choose  $\rho = 0.1 \text{ km}^{-1}$  (i.e. 10 km) which is the average distance travelled by labour force (statistics Denmark [23]). To allow some long-distance interactions, we introduced a parameter  $\epsilon_p = 4\%$  representing the fraction of distance-independent connections. This value we based on the fraction of workers travelling longer than 50 km to work (statistics Denmark [23]).

### 3.1.2. Fits and predictions

We defined an early phase to be the period of time when between 0.1% and 1% of the population were infected (blue lines figure 3a). We then fitted  $\beta$  and a time delay,  $\tau$ , to the SEIR model with a  $\chi^2$ -fit (assuming Poissonian statistics) and kept  $\lambda_E$  and  $\lambda_I$  fixed to the true numbers (used in the simulation). The initial number of infected,  $N_{\text{init}}$ , was also fixed to the true numbers. The fit parameters were then inserted into the SEIR model, and  $I_{\text{peak}}^{\text{fit}}$  and  $R_{\infty}^{\text{fit}}$  were extracted from the fitted model and compared to the  $I_{\text{peak}}^{\text{ABM}}$  and  $R_{\infty}^{\text{ABM}}$  from the ABM simulation.

Data accessibility. Data and relevant code for this research work are stored in GitHub: [www.github.com/ChristianMichelsen/NetworkSIR](http://www.github.com/ChristianMichelsen/NetworkSIR) and have been archived within the Zenodo repository: <https://zenodo.org/badge/latestdoi/258223118>.

Authors' contributions. C.M.: conceptualization, formal analysis, investigation, methodology, software, validation, visualization, writing—original draft, writing—review and editing; E.S.M.: investigation, software, validation, visualization, writing—review and editing; L.E.C.: supervision, validation, writing—review and editing; T.C.P.: conceptualization, investigation, methodology, project administration, software, supervision, validation, visualization, writing—original draft, writing—review and editing; M.L.H.: conceptualization, formal analysis, investigation, methodology, software, validation, visualization, writing—original draft, writing—review and editing; M.H.J.: formal analysis, investigation, supervision, validation, writing—review and editing; T.H.: conceptualization, investigation, supervision, validation, visualization, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein. Conflict of interest declaration. We declare that we have no competing interests.

Funding. M.L.H. acknowledges the Carlsberg Foundation grant no. CF20-0621 and the Lundbeck Foundation grant no. R347-2020-2250. E.S.M. and M.H.J. acknowledge support from the Independent Research Fund Denmark grant no. 9040-00116B and Danish National Research Foundation through StemPhys Center of Excellence, grant no. DNRF116.

Acknowledgements. The authors are grateful to the Danish expert group of SARS-CoV-2 modelling led by Statens Serum Institute, especially Robert L. Skov, Kåre Mølbak, Camilla Holten Møller, Viggo Andreassen, Kaare Græsbøl, Theis Lange, Carsten Kirkeby, Frederik P. Lyngse, Matt Denwood, Jonas Juul, Sune Lehman, Uffe Thygesen and Laust Hvas Mortensen. Furthermore, we thank Kim Sneppen for valuable discussions.

## References

- Chinazzi M *et al.* 2020 The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400. (doi:10.1126/science.aba9757)
- WHO: see [www.who.int/news-room/detail/27-04-2020-who-timeline-covid-19](http://www.who.int/news-room/detail/27-04-2020-who-timeline-covid-19) (accessed 29 September 2020).
- Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. 2020 How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* **395**, 931–934. (doi:10.1016/S0140-6736(20)30567-5)
- Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Flasche S. 2020 Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Global Health* **8**, e488–e496. (doi:10.1016/S2214-109X(20)30074-7)
- Keeling MJ, Hollingsworth TD, Read JM. 2020 Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). *J. Epidemiol. Commun. Health* **74**, 861–866. (doi:10.1101/2020.02.14.20023036)
- Kuniya T. 2020 Prediction of the epidemic peak of coronavirus disease in Japan, 2020. *J. Clin. Med.* **9**, 789. (doi:10.3390/jcm9030789)
- Ghafouri-Fard S, Mohammad-Rahimi H, Motie P, Minabi MA, Taheri M, Nateghinia S. 2021 Application of machine learning in the prediction of COVID-19 daily new cases: a scoping review. *Heliyon* **7**, e08143. (doi:10.1016/j.heliyon.2021.e08143)
- Fokas AS, Dikaios N, Katsis GA. 2020 Mathematical models and deep learning for predicting the number of individuals reported to be infected with SARS-CoV-2. *J. R. Soc. Interface* **17**, 20200494. (doi:10.1098/rsif.2020.0494)
- Kermack WO, McKendrick AG. 1927 A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* **115**, 700–721. (doi:10.1098/rspa.1927.0118)
- Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. 2020 Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493. (doi:10.1126/science.abb3221)
- Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Abbott S. 2020 The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* **5**, e261–e270. (doi:10.1016/S2468-2667(20)30073-6)
- van Bunnik BA, Morgan AL, Bessell P, Calder-Gerver G, Zhang F, Haynes S, Lepper HC. 2020 Segmentation and shielding of the most vulnerable members of the population as elements of an exit strategy from COVID-19 lockdown. *medRxiv* (doi:10.1101/2020.05.04.20090597).
- Danon L, Brooks-Pollock E, Bailey M, Keeling MJ. 2020 A spatial model of COVID-19 transmission in England and Wales: early spread and peak timing. *medRxiv* (doi:10.1101/2020.02.12.20022566).
- Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. 2020 Modelling transmission and control of the COVID-19 pandemic in Australia. *Nat. Commun.* **11**, 1–13. (doi:10.1038/s41467-020-19393-6)
- Sneppen K, Nielsen BF, Taylor RJ, Simonsen L. 2021 Overdispersion in COVID-19 increases the effectiveness of limiting nonrepetitive contacts for transmission control. *Proc. Natl Acad. Sci. USA* **118**, e2016623118. (doi:10.1073/pnas.2016623118)
- Milne GJ, Xie S. 2020 The effectiveness of social distancing in mitigating COVID-19 spread: a

- modelling analysis. *medRxiv* (doi:10.1101/2020.03.20.20040055).
17. Bansal S, Grenfell BT, Meyers LA. 2007 When individual behaviour matters: homogeneous and network models in epidemiology. *J. R. Soc. Interface* **4**, 879–891. (doi:10.1098/rsif.2007.1100)
  18. Kong L, Wang J, Han W, Cao Z. 2016 Modeling heterogeneity in direct infectious disease transmission in a compartmental model. *Int. J. Environ. Res. Public Health* **13**, 253. (doi:10.3390/ijerph13030253)
  19. Kang D, Choi H, Kim JH, Choi J. 2020 Spatial epidemic dynamics of the COVID–19 outbreak in China. *Int. J. Infect. Dis.* **94**, 96–102. (doi:10.1016/j.ijid.2020.03.076)
  20. Giuliani D, Dickson MM, Espa G, Santi F. 2020 Modelling and predicting the spatio-temporal spread of COVID-19 in Italy. *BMC Infect. Dis.* **20**, 1–10. (doi:10.1186/s12879-020-05415-7)
  21. Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. 2019 Complexity of the basic reproduction number (R0). *Emerg. Infect. Dis.* **25**, 1–4. (doi:10.3201/eid2501.171901)
  22. Danish Serum Institute: www.ssi.dk (accessed 29 September 2020).
  23. Statistics Denmark: www.statistikbanken.dk (accessed 29 September 2020).
  24. Wong DW, Li Y. 2020 Spreading of COVID-19: density matters. *PLoS ONE* **15**, e0242398. (doi:10.1371/journal.pone.0242398)
  25. Ganasegeran K, Jamil MFA, Ch'ng ASH, Looi I, Peariasamy KM. 2021 Influence of population density for COVID-19 spread in Malaysia: an ecological study. *Int. J. Environ. Res. Public Health* **18**, 9866. (doi:10.3390/ijerph18189866)
  26. Kodera S, Rashed EA, Hirata A. 2020 Correlation between COVID-19 morbidity and mortality rates in Japan and local population density, temperature, and absolute humidity. *Int. J. Environ. Res. Public Health* **17**, 5477. (doi:10.3390/ijerph17155477)
  27. Bhadra A, Mukherjee A, Sarkar K. 2021 Impact of population density on COVID-19 infected and mortality rate in India. *Model. Earth Syst. Environ.* **7**, 623–629. (doi:10.1007/s40808-020-00984-7)
  28. Chen K, Li Z. 2020 The spread rate of SARS-CoV-2 is strongly associated with population density. *J. Travel Med.* **27**, taaa186. (doi:10.1093/jtm/taaa186)
  29. Martins-Filho PR. 2021 Relationship between population density and COVID-19 incidence and mortality estimates: a county-level analysis. *J. Infect. Public Health* **14**, 1087–1088. (doi:10.1016/j.jiph.2021.06.018)
  30. Hittner JB, Fasina FO, Hoogesteijn AL, Piccinini R, Maciorowski D, Kempaiah P, Rivas AL. 2021 Testing-related and geo-demographic indicators strongly predict COVID-19 deaths in the united states during March of 2020. *Biomed. Environ. Sci.* **34**, 734–738.
  31. Huang S, Li J, Dai C, Tie Z, Xu J, Xiong X, Lu C. 2021 Incubation period of coronavirus disease 2019: new implications for intervention and control. *Int. J. Environ. Health Res.* **32**, 1707–1715. (doi:10.1080/09603123.2021.1905781)
  32. Gillespie DT. 1977 Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361. (doi:10.1021/j100540a008)
  33. Boligsiden: www.boligsiden.dk (accessed 29 September 2020).
  34. HOPE project: www.hope-project.dk (accessed 29 September 2020).
  35. Andreasen V, Viboud C, Simonsen L. 2008 Epidemiologic characterization of the 1918 influenza pandemic summer wave in Copenhagen: implications for pandemic control strategies. *J. Infect. Dis.* **197**, 270–278. (doi:10.1086/524065)
  36. Arcede JP, Caga-Anan RL, Mentuda CQ, Mammeri Y. 2020 Accounting for symptomatic and asymptomatic in a SEIR-type model of COVID-19. *Math. Model. Nat. Phenomena* **15**, 34. (doi:10.1051/mmnp/2020021)
  37. Guan J, Zhao Y, Wei Y, Shen S, You D, Zhang R, Chen F. 2022 Transmission dynamics model and the coronavirus disease 2019 epidemic: applications and challenges. *Med. Rev.* **2**, 89–109. (doi:10.1515/mr-2021-0022)
  38. Holmdahl I, Buckee C. 2020 Wrong but useful—what COVID–19 epidemiologic models can and cannot tell us. *N. Engl. J. Med.* **383**, 303–305. (doi:10.1056/NEJMp2016822)
  39. Wynants L, Van Calster B, Bonten MM, Collins GS, Debray TP, De Vos M, Schuit E. 2020 Prediction models for diagnosis and prognosis of COVID–19 infection: systematic review and critical appraisal. *BMJ* **369**, m1328. (doi:10.1136/bmj.m1328)
  40. Rivas AL, Fasina FO, Hoogesteyn AL, Konah SN, Febles JL, Perkins DJ, Smith SD. 2012 Connecting network properties of rapidly disseminating epizoonotics. *PLoS ONE* **7**, e39778. (doi:10.1371/journal.pone.0039778)