Machine learning spatio-temporal epidemiological model to evaluate Germany-county-level COVID-19 risk

IOP Publishing Mach. Learn.: Sci. Technol. 2 (2021) 035031 https://doi.org/10.1088/2632-2153/ac0314

MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
7 March 2021

REVISED
8 May 2021

ACCEPTED FOR PUBLICATION
19 May 2021

PUBLISHED
13 July 2021

PAPER

Machine learning spatio-temporal epidemiological model to evaluate Germany-county-level COVID-19 risk

Lingxiao Wang[1], **Tian Xu**[2], **Till Stoecker**[3], **Horst Stoecker**[1,4], **Yin Jiang**[2,5,*] and **Kai Zhou**[1,*]

[1] Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany
[2] Department of Physics, Beihang University, Beijing 100191, People's Republic of China
[3] Black Hole KG, 61440 Oberursel (Taunus), Germany
[4] Institute for Theoretical Physics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany
[5] Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine and Engineering, Beihang University, Beijing 100191, People's Republic of China
[*] Authors to whom any correspondence should be addressed.

E-mail: jiang_y@buaa.edu.cn **and** zhou@fias.uni-frankfurt.de

Keywords: SUIR model, COVID-19 pandemic, deep learning, cellular automata

Abstract

As the COVID-19 pandemic continues to ravage the world, it is critical to assess the COVID-19 risk timely on multi-scale. To implement it and evaluate the public health policies, we develop a machine learning assisted framework to predict epidemic dynamics from the reported infection data. It contains a county-level spatio-temporal epidemiological model, which combines spatial cellular automata (CA) with time sensitive-undiagnosed-infected-removed (SUIR) model, and is compatible with the existing risk prediction models. The CA-SUIR model shows the multi-scale risk to the public and reveals the transmission modes of coronavirus in different scenarios. Through transfer learning, this new toolbox is used to predict the prevalence of multi-scale COVID-19 in all 412 counties in Germany. A t-day-ahead risk forecast as well as assessment of the non-pharmaceutical intervention policies is presented. We analyzed the situation at Christmas of 2020, and found that the most serious death toll could be 34.5. However, effective policy could control it below 21thousand, which provides a quantitative basis for evaluating the public policies implemented by the government. Such intervening evaluation process would help to improve public health policies and restart the economy appropriately in pandemics.

1. Introduction

Public health safety is the cornerstone of economic and social stability. The CoronaVirus Disease 2019 (COVID-19) is an infectious disease caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) which became a global pandemic that spread across the world in 2020. As of early March 2020, the number of reported cases in Germany experienced a rapid growth. Only severe restrictions on contacts and travel can ease the situation in the summer. However, the number of infected cases have rapidly increased again since mid-September, named popularly as the second-wave pandemic worldwide. By January 25, 2021, the pandemic has caused 2141 665 confirmed cases and 52 087 fatalities in Germany[6]. Many states are facing the dilemma challenges of public heath and economics in curbing the spread of the COVID-19. The medical resources of Germany (ICU beds, health workers and doctors) have not arrived at their ceiling, but the current phase of the pandemic is not encouraging. As a highly-infectious disease, COVID-19 is expected to remain spreading and be even worse to cause a higher number of infections and fatalities in the future. No well adapted medical treatments nor a sufficient vaccine allocation is available at this moment, so the non-pharmaceutical public health policies, such as, social distancing and travel restrictions can be implemented to alleviate the spread of epidemic currently. This policy has proven an effective containment in

[6] Statistics is from the Coronavirus Disease 2019 (COVID-19)—Situation Report of the Robert Koch Institute (RKI).

East Asia, e.g. China [1], South Korea and Japan. In central Europe this strategy worked as well at early stage [2]. In the presently worsening health crisis, the need to establish a multi-scale (nation-state-county) COVID-19 risk prediction model emerges, which can guide the evaluation and improvement on the public policies of governments (states) at the different regional levels. The multi-scale risk information can help to improve the non-pharmaceutical intervention strategies such as cross-state travel restriction, home isolating regulations and allocation of vaccination resources. The public can also benefit from the information, to make effective access of medical resources and help implementing or following the public policies effectively.

Evaluation of the risk of the on-going COVID-19 pandemic across the diverse geographical administrative units is challenging, because of the substantial heterogeneity rooted in the federal state political system with diverse economic composition, traffic conditions and individual political views. Our research presents the needed tool which can optimize our understanding of the importance of the local and regional (state-county-level) pandemic dynamics. Studies devoted to the dynamics of the epidemics based on mathematical models [3, 4] (also see reviews [5–7]) have been mainly focusing on time-domain analysis and ignored the influence of geographic distribution and the population migration [2], which in fact presents cross-county transfers on a daily-scale. A model which can predict the details of the spatio-temporal evolution of the COVID-19 cases is urgently needed to appropriately assess the local risk in a county of politically diverse multi-scale regions. Works have been concerned with spatial distribution or local evolutions were mainly based on historical statistical data [8–10]. Several recent studies share similar goals. Different countries [11–14] merited this as an urgent need across the world. From a practical perspective, an intellectual and efficient risk prediction model needs to be performed on a county basis, like with all Landkreise[7] in Germany, since people are highly mobile and connective within Germany and to all of Europe as well as all other continents.

Machine learning (ML), a branch of artificial intelligence (AI), efficiently integrates statistical and inference algorithms and thus offers the opportunity to uncover hidden structures of evolution in complex data and to describe it with finite dynamical parameters. Therefore, a combination of ML algorithms and spatio-temporal epidemiological models for risk prediction in the context of COVID-19 pandemic is a promising option. Although the state-of-the-art ML techniques have been deployed into applications of COVID-19 pandemic predictions [15–18], the highest concern for the governments and residents is how fast the COVID-19 spreads. Available ML models lack of capability of predicting the needed multi-scale evolution, since they routinely focus on this exhibit promising implications for individuals and on a national scale, and thus are impeded by the paucity of validation information and limiting privacy rules in Germany.

This study aims to establish such a new evaluation paradigm by introducing a spatio-temporal ML assisted epidemiological model. This dynamical model consists of cellular automaton (CA) [19, 20] plus a modified SUIR model [13, 21], where the dynamical parameters are extracted from the reported data of Robert Koch-Institut (RKI) by the Conv2LSTM Neural Network, a deep learning technique regularly applied in video analysis. The present SUIR model is more inclusive than previously used susceptible-infectious-recovery (SIR) models, as it considers explicitly the Undetected or Asymptomatic populations [22, 23], hidden due to the lack of testing and also the systematic uncertainties for the testing. We reproduces the evolution of infection maps from 25 February to 28 March 2020, and extracts the infectious parameters from RKI data in September 2020 with a neural network which was training on the data-set from CA-SUIR model simulations. The well-learned machine predicts county-level infection maps from September to November 2020, which is matched with the reported situation. With regard to the high-profile holiday, the different public health policies are evaluated in the end of the section, in which the entire, partial and none lock-down strategies are compared.

2. Machine learning spatio-temporal epidemiological model

The full flow chart is assembled in figure 1. to present the multi-scale spatio-temporal simulation of epidemics, where as the first step real maps are projected onto the topological equivalent maps as explained in figure 8. The cumulative infected cases in the real profiles[8] are reassembled into a topological equivalent map with 25×25 sites in Germany, which contains 412 counties (see details in figure 8). The micro-states of each county, its number of infected cases, are driven by the SUIR and CA simultaneously in the model, which construct the infection map. With simulated data from the CA-SUIR model, the Conv2LSTM neural networks [24] are designed to learn to extract dynamical parameters of CA-SUIR model. Subsequently, we transfer the well-trained neural networks to make timely prediction and risk evaluation on RKI infection maps.

[7] It is the primary administrative subdivision higher than a city in Germany, also could be named as county in English.
[8] See the map at COVID-19 pandemic in Germany.

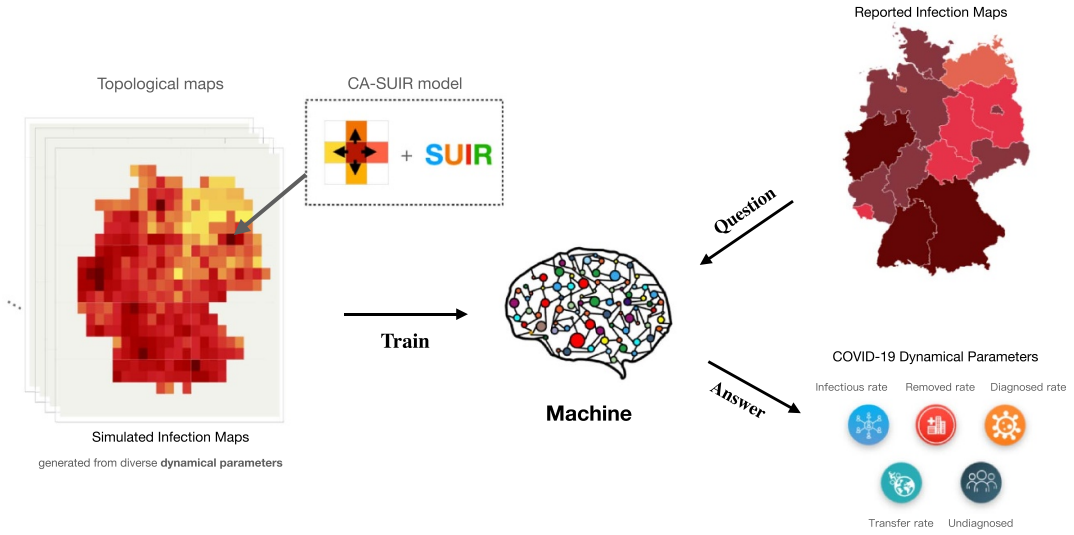

Figure 1. A framework to evaluate the COVID-19 risk in Germany. The left panel is the simulated data-set of the infection cases in Germany, which is generated from diverse dynamical parameters in cellular automata (CA)-sensitive-undiagnosed-infected-removed (SUIR) model. The simulated data-set is fed to the central part, a machine designed as the Conv2LSTM neural networks. In the transfer learning manner, the well-trained machine is deployed to extract the dynamical parameters of COVID-19 pandemic from reported infection data map.

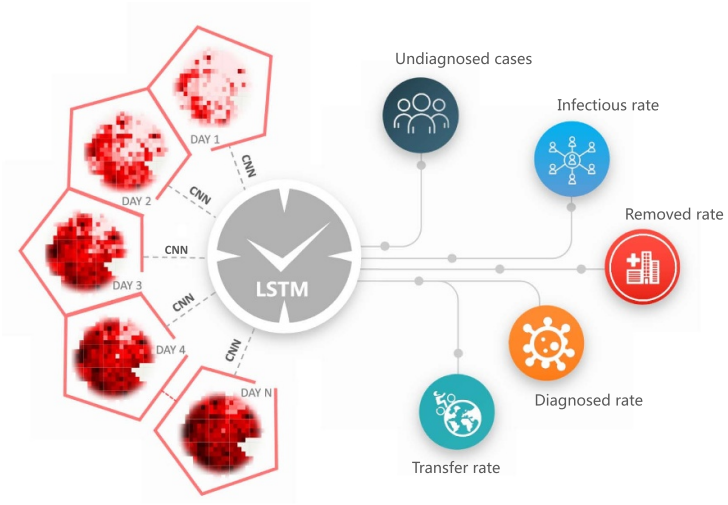

Figure 2. Machine learning the dynamical parameters of CA-SUIR model from the evolution of infection maps. The CNNs layers and LSTM units are combined to be trained on the inputs which are short movies. The undiagnosed population, infectious rate, removed rate, diagnosed rate and transfer rate are decoded from the deep neural networks.

2.1. Extracting dynamical parameters via deep learning

To prepare training data for the Conv2LSTM, we generate 30-day simulation maps as short movies by associating the SUIR model parameters in a 10% uncertainty range. The total data size is 10 000, in which 90% is treated as the training data-set and the rest is the testing data-set. The 30 day movie has 30 frames which is fed to the machine with continuous 7 day segments as slip windows [25]. The structure of the Conv2LSTM neural networks are exhibited in figure 10. The Conv2LSTM neural networks we built are demonstrated in figure 2, in which the time distributed convolutional neural networks (CNNs) layers and long short-term memory (LSTM) layers are combined to decode the information hidden in the short video. The details of the structure are arranged in figure 10. The time distributed CNNs could extract features (with temporal structure preserved) from sequential image inputs and then embed them into the following LSTM layers, which is able to efficiently drill evolution rules from the short video.

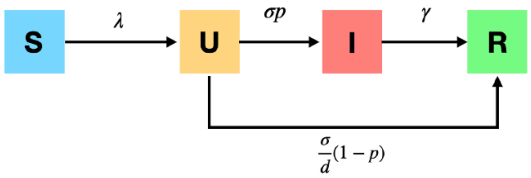

Figure 3. The SUIR diagram shows how individuals are infected and transferred among each compartments in a close area.

2.2. Spatio-temporal CA-SUIR model

We construct a cellular automaton model for tracking the epidemics spatial evolution in a closed system (nation with fixed boundary as figure 1 shown). The underlying structure is a $L \times L$ cell grid, where L is the system size, and the number is from the topological map. The color of cell characterizes the population in minimum geographical unit, which is county in our case. It is an instructive example to set the cell size to a typical area of city, such as 10^2–10^3 km^2, which can simulate the epidemics in the corresponding resolution approximately. The resolution could be naturally increased into community level, but it needs more detailed data which is laboriously accessible, thus the following discussion focuses on county level and corresponding resolution. This model adopts the Moore neighbor, and cells update their states by transition tensor $P^i_{m,n}(t)$ with $4 \times L \times L$ shape, which means the possibility of transfer from city (m, n) to direction i at t time-step. The $P^i(t)$ are 4 matrices $(P^{\mathrm{u}}(t), P^{\mathrm{d}}(t), P^{\mathrm{l}}(t), P^{\mathrm{r}}(t))$ with same lattice size, in which different colors labels the direction of transfer. In the following simulation, we adopt the mean field approximation [26], which leads to the uniform transition possibility $P^{\mathrm{u}}(t) = P^{\mathrm{d}}(t) = P^{\mathrm{l}}(t) = P^{\mathrm{r}}(t) = \epsilon$. The transfer rate ε is introduced to describe the average movement of the residents on each site. To mimic the border control induced closed boundary, a transfer rate $\varepsilon = 0$ has been used in the primary epidemic stage [27]. The evolution of the CA-SUIR model is as follows: Initialization. Set the position of infection burst (x, y) and generate population distribution $S(t)$ at the $L \times L$ lattice. At $t = 0$ time-step, epidemic turns on and population begins to migrate; SUIR dynamics. At t time-step, evolve SUIR model t' times on each site; movement. The parts of Undiagnosed people on site (m, n) migrate into the neighbor site with transition matrices $P^k(t, m, n)$ at $t + 1$ time-step; update all sites synchronously until T time-step is reached.

The temporal evolution parts are driven by the SUIR model and are demonstrated in figure 3, where the infectious rate λ controls the rate of spread which represents the probability of transmitting disease between a susceptible and an infectious individual. The diagnosed rate σ dictate the probability of latent infectious individuals becoming confirmed by testing. Removed rate γ contains the rate of mortality and recovery from the infectious. Considering the special features of COVID-19, we introduce a considerable population that could be infected but without any symptoms, or, they just have no access to perform a detection(with possibility p), which is concluded in a new compartment, undiagnosed (U). In contrast to the infected who is usually subjected to be isolated in medical facilities or quarantine at home, the undiagnosed population has normal mobility which leads to the growing-up infection; the infectivity could decrease after d days [28], eventually removed from the infection population, which means immunity will retain at least several months [29]. In a closed population with no birth or mortality per day, the modified SUIR model becomes,

$$\frac{dS}{dt} = -\frac{S}{N}\lambda U \tag{1}$$

$$\frac{dU}{dt} = \frac{S}{N}\lambda U - \sigma p U - \frac{\sigma(1-p)}{d}U, \tag{2}$$

$$\frac{dI}{dt} = \sigma p U - \gamma I \tag{3}$$

$$\frac{dR}{dt} = \gamma I + \frac{\sigma(1-p)}{d}U, \tag{4}$$

where $N = S + U + I + R$ is the total population and we adopt a redefinition as $\sigma \equiv \sigma p + \frac{\sigma(1-p)}{d}$ in the results.

It should be noted that the choice of the spatial or temporal models are not unique. The modified SIR (or SEIR) models are routinely adopt in forecasting COVID-19 temporal evolution. Although it is not able to properly forecast the long-term situation [30], it offers a reasonable base-line for short-term prediction

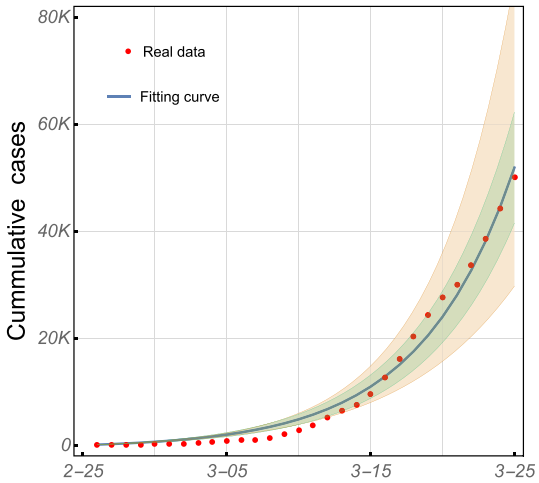

Figure 4. The comparison of RKI data and simulated cumulative infection cases. The data is from 25 February 2020 to 28 March 2020, in which the shadow bars are from uncertainty estimation. The orange panel is the variation space of SUIR model parameters, which shows a reliable area under 10% changes. The green panel is from the 10% uncertainty of the undiagnosed population (U).

which could be used to train the deep learning machine sufficiently. On the other side, the temporal model could be improved by introducing more realistic factors, such as the risk of reinfection [31], the isolation measures for patients with mild symptoms [32] or the close contacts [33]. With regard to the spatial part, the philosophy is similar. The accurate tracking of the individuals [34] or agent-based models [8] could improve the performance of the prediction on dynamical transmission of COVID-19 in Germany.

3. Results

3.1. Pandemic dynamics modeling, data-set generation and network capacity

With the daily number of cumulative cases in Germany (Data is from the open database of RKI [9]), we determine the temporal parameters of the CA-SUIR model to be infectious rate $\lambda = 0.293$, diagnosed rate $\sigma = 0.129$, and removed rate $\gamma = 0.628$ by matching well with the first-wave burst out in March especially the rapidly increasing stage (from 25 February 2020 to 28 March 2020), as presented in figure 4. The initial undiagnosed cases remain same spatial distribution as the reported infection cases with $U_0 \simeq 700$, meanwhile the transfer rate connected counties is set as $\varepsilon = 0.08$. Figure 4 suggests that there is an overestimation of the simulation at early stage (first 20 days), which could be naturally understood to be related with the underestimation in reality at that time since the limited detection ability and the supply of test reagents were insufficient [35]. The estimation to the uncertainty of undiagnosed population shown in figure 4, which is actually from the stability issues of RT-PCR testing [36] and unreported cases in practical situation.

We train the Conv2LSTM to predict the dynamical parameters of the involved epidemiological model (see the learning curve in figure 11). The four sub-figures in figure 5 demonstrate the comparison between the prediction from the Conv2LSTM and ground-truth of testing data, they are the transfer rate ε, the infectious rate λ, the diagnosed rate σ and undiagnosed population U respectively. The well-trained machine extracts parameters from the RKI infection map of March 2020 as $\lambda = 0.26, \sigma = 0.09, U_0 = 1352, \epsilon = 0.068$. They are numerically matched with the fitting results, in which the diagnosed rate and initial undiagnosed population maintain the consistency with fitting parameters in the relationship σU_0.

3.2. Machine learning-assisted risk prediction

To assess the COVID-19 risk in realistic situation timely, we deploy the well-trained neural network into the current cases in transfer learning manner. Transfer learning is a ML method where a model developed for a task is generally reused as the starting point for a model on a second task. In our case, we transfer the pre-trained models into the on-going second-wave COVID-19 epidemics in Germany, extracting the dynamical parameters of the spatio-temporal pandemics from the 30 day RKI infection maps with beginning from September 8, 2020, which yields the dynamical parameters as $\lambda = 0.21, \sigma = 0.167, U_0 = 7426, \epsilon = 0.09$.

Figure 6 describes the homogeneity between simulation and RKI infection maps from September to November 2020 in Germany. The former cumulative infection maps shown in figure 6(a) are generated from

[9] RKI COVID-19 and COVID-19 GeoHub Deutschland.

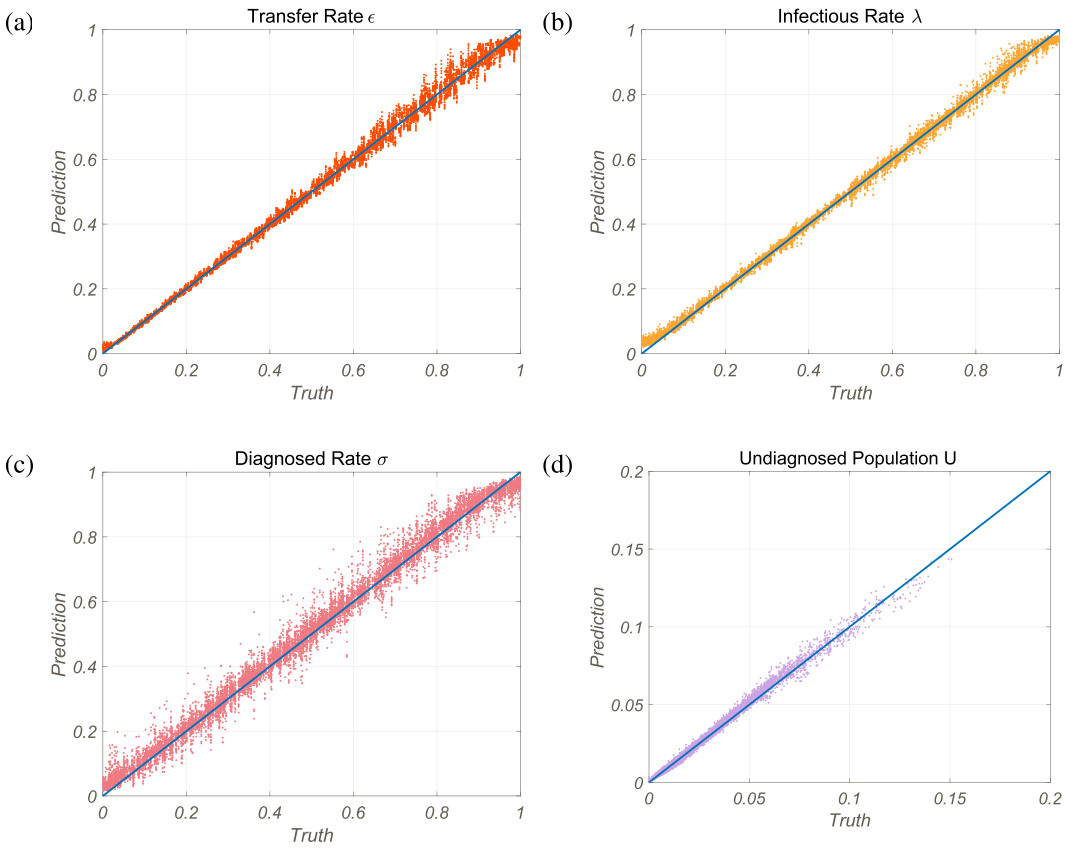

Figure 5. The performance of Conv2LSTM neural networks to extract dynamical parameters on testing data-set. The parameters are rescaled in $(0, 1)$ interval, in which the undiagnosed population U only shows $(0, 0.2)$ since the large numbers are few. They manifest high coefficient of determination as $R^2(\varepsilon) = 0.996$, $R^2(\lambda) = 0.998$, $R^2(\sigma) = 0.987$, $R^2(U) = 0.993$.

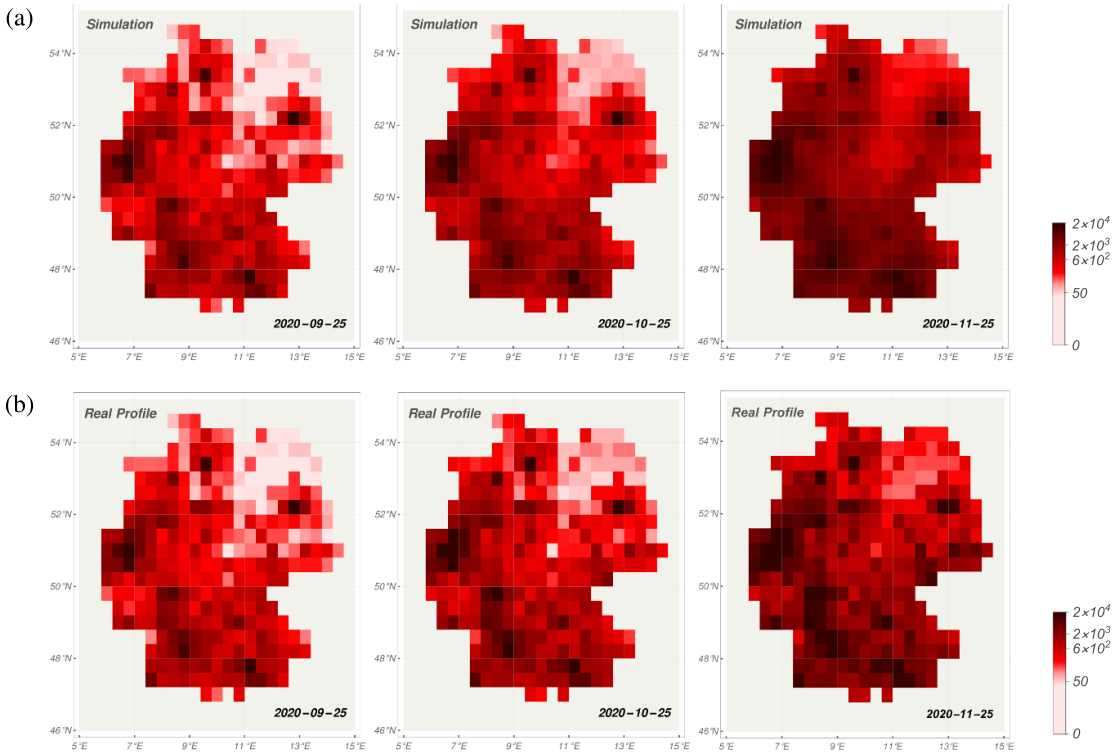

Figure 6. Comparison of simulation and RKI infection maps from September to November 2020 in Germany. The color bar is set into an uniform red color scale, the darkest color is 20 000 cases with a logarithmic re-scale. The cumulative infection cases are presented at 2020-09-25, 2020-10-25 and 2020-11-25 respectively, whose dynamical parameters are extracted from the Conv2LSTM. The real profiles show in down panel.

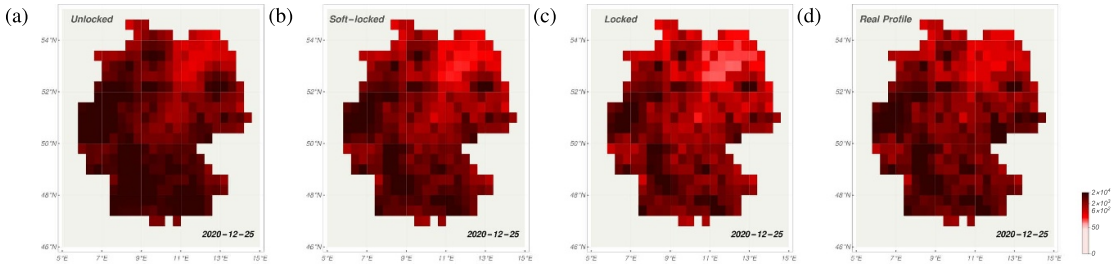

Figure 7. Cumulative infection maps at Christmas of 2020. The color bar is set into an uniform red color scale, the darkest color is 20 000 cases with a logarithmic re-scale. The dynamical parameters are extracted from RKI infection maps by the well-trained machine. (a) The result is based on the existing public policies. (b) The soft lockdown rules contains the relative loose restrictions to some hot spots, thus the diffusion of the epidemic is alleviated partially. (c) In locked, the strictest rules prohibit any form of mobility, which yield the least infections. (d) The reported cumulative infection maps from RKI data.

the SUIR model with the ML parameters during 3 months. The latter maps shown in figure 6(b) are projected from the RKI data, in which we present the cumulative infection cases in real profiles until the end of November. The top 5 biggest cities, Berlin (13°23E, 52°31N), Hamburg (10°00E, 53°33N), München (11°35E, 48°03N), Köln (6°58E, 50°56N) and Frankfurt (8°41E, 50°07N) demonstrate a worse situation than the other sites. The slight mismatching on the map of November is due to the actual soft lock-down policy has been executed from 2 November. The functional regulations will contain the pandemic moderately, which will be examined under different policies in the following part.

3.3. Public policies evaluation

In public policy, three different degrees of restriction rules, named as the unlocked, soft-locked and locked strategies are evaluated. The last one is the lock-down policy came into play first in Wuhan [37, 38] which shows strictest restriction to the mobility of residents compared with the other policies. The restrictive rules take effect in different places with differentiated implementation, which includes but are not limited to office, campus, school, restaurant and cross-county travel. How is the effect of implementation? It is of crucial concern for both governments and residents who are enduring the social-economical pressure in pandemic [11, 39–41]. The comprehensive influences at county-level are averaged as an equivalent effect to the transfer rate, which means the effect of policies are coarse graining to the county-level. Reasonably, it could be improved if we collect more precise data below county-level, e.g. individual or community-level [42].

The differences among them in simulation are the transfer rate and susceptible population: the most strict locked rules has transfer rate $\varepsilon = 0$, the soft-locked with half transfer $\varepsilon = 0.045$, and the rule with no additional restriction has $\varepsilon = 0.09$ as default; meanwhile, the containment policies could result in a consequence of infectious population decreasing that effectively deplete the susceptible population [1]. It leads to the reduction of the susceptible population, reaching to 1/2, 1/8 for the locked and soft-locked rules respectively, in which the 1/8 is from the rough estimation to the necessary out time. These three strategies are implemented from the same time point 25 November 2020, and their 30 day simulation results are shown in figure 7. In our evaluation to the policies, 2.3 million people would be infected at Christmas of 2020 in the worst situation, where daily increase is approximately 80 thousand, which means the total mortality will reach 34.5 thousand approximately (estimate with mortality rate ~1.5% [43]). With regard to the soft-locked policy, the infection number reduces to 1.4 million with 21 thousand fatalities and the corresponding daily increasing cases are 20 thousand. In the most strict rules, although it restricts the mobility, the number of infection cases will be 1.07 million and the daily increase is only 4 thousand. Compared with the reported cases shown in figure 7(d), they achieve the consistency in geographic distribution. Considering the reported cases from RKI, cumulative cases 1612 648 with daily increasing 25 533 and 29 182 fatalities, the lock-down policies have clearly made a difference. It can effectively reduce the chance of exposure to the infectious population, which reflects in the reduction of the transfer rate and effective susceptible population in our model. It is also found that the efficiency of the vaccination or allocation of the vaccine could be embedded into our CA-SUIR model through the similar method that reduction to the susceptible population [1, 44], which is mathematically equivalent with introducing a suppression factor to the susceptible population as $-\nu S$. The factor ν could include a time dependent effect to show the progress of the vaccination in different counties. As a matter of fact, our framework has the alignment with reality which suggests a succinct scheme for quantifying the effects of the public policies. The quantification of policy effects is conceivably shown in the variation of the dynamical parameters or populations.

4. Discussion

In this paper, we construct a multi-scale spatio-temporal epidemiological model, which enables the prediction on the state-county-level COVID-19 risk over 412 Landkreise in Germany. Reasonably, it is convenient to replace the module by other elaborate spatial or temporal models for reaching a more accurate prediction in our framework. Regarding the ML module, transfer learning helps us to extract the dynamical parameters from the RKI infection maps through the well-trained Conv2LSTM neural networks.

The CA-SUIR model reproduces the epidemics in March and September acknowledged as the first- and second-wave of COVID-19 pandemic. Even though the prediction focuses on the infectious dynamics in the current paper, it is conveniently feasible to derive recovery and mortality rate from the model under the guarantee of accurate data sources confronted. The simulations perform properly to describe and predict the data in the initial 30 day phase, while it shows a tendency towards a faster smoothing-out of the pronounced local fine structure and persistent hot spots, as compared to the data. It could be understood that the observed infection cases concentrated in the big cities, such as Berlin, Hamburg, München, Köln and Frankfurt than the other sites in the simulation, which reflects high population density enhances the diffusion of the COVID-19. The prediction could be evidently improved by introducing more abundant interactions across and within the counties. Some practical improvements can be implemented in future works, such as replacing the geographic lattice to traffic networks [5, 42, 45–47], introducing the higher resolution lattices beyond the county-level which also means higher-order interactions among the residents [12, 20, 42, 48]. The well-trained machine generates the infection maps which are consistent with the RKI data in 3 months. With regard to the slight uncertainty in extracting dynamical parameters, it could be avoided by training the neural networks on more heterogeneous data-set. We will present the further improvement in our future papers.

The ML-assisted CA-SUIR model indicates that knowledge of risk evaluation in the county-level can be exhibited to the public directly and timely, meanwhile, the 7 day windows could forecast the current situation robustly. However, we could not ignore the limitation of this study. The main shortage of our framework is highly dependent on the accuracy of the CA-SUIR model which is oversimplified for the complicated COVID-19 situation. Our study was also limited by the absence of accurate mobility data which unavoidably affects the feasibility. The assessment to the different public health policies suggests that the transmission modes of coronavirus could be shaped by the non-pharmaceutical interventions dramatically. The other valuable application of our evaluation system is to improve allocation of medical resources which are usually unequal in different counties, some of them are suffering from the epidemic that far exceeds their medical capacity.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://covid-19-geohub-deutschland-esridech.hub.arcgis.com/search?collection=Dataset and https://github.com/Anguswlx/Conv2LSTM-COVID19.

Acknowledgments

The authors thank Esteban Vargas, Maria Babarossa and Jan Fuhrmann for useful discussions and comments. The work on this research is supported by the BMBF through the ErUM-Data funding and the Samson AG AI Grant (L W, K Z), by the Walter Greiner Gesellschaft zur Foerderung der physikalischen Grundlagenforschung e.V. through the Judah M. Eisenberg Laureatus Chair at Geothe Universitaet Frankfurt am Main (H S), by the National Natural Science Foundation of China, Grant No. 11875002 (T X, Y J), by the Zhuobai and Bajian Programs of Beihang University (Y J). The authors also thank the donation of NVIDIA GPUs from NVIDIA Corporation.

Appendix A. Projection of infection map

In order to simulate the population evolution on lattice, the topological mapping is adopted in our CA computation. We firstly embed the map in a rectangle area whose length and width are chosen as the maximum value of the Germany map, as in figure 8(left). Then the area is segmented into uniform sites of $L \times L$. And each site is labelled by its row and column numbers as B_{mn}, where $1 \leqslant m, n \leqslant L$. If a county occupies N sites, all kinds of the population, such as susceptible, undiagnosed, infected and removed, are

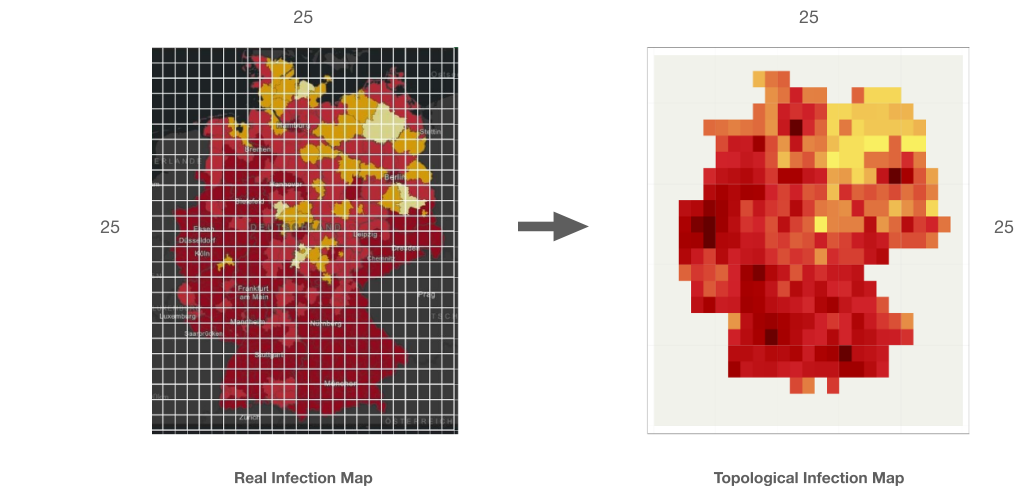

Figure 8. Mapping from real infection map of Germany onto topological equivalent lattice. The left map is divided into 25 × 25 sites as the size of the right part. The geometric shape of the square lattice is irrelevant to the real geographic topography. The site on lattice depict a typical county-level unit of the real map.

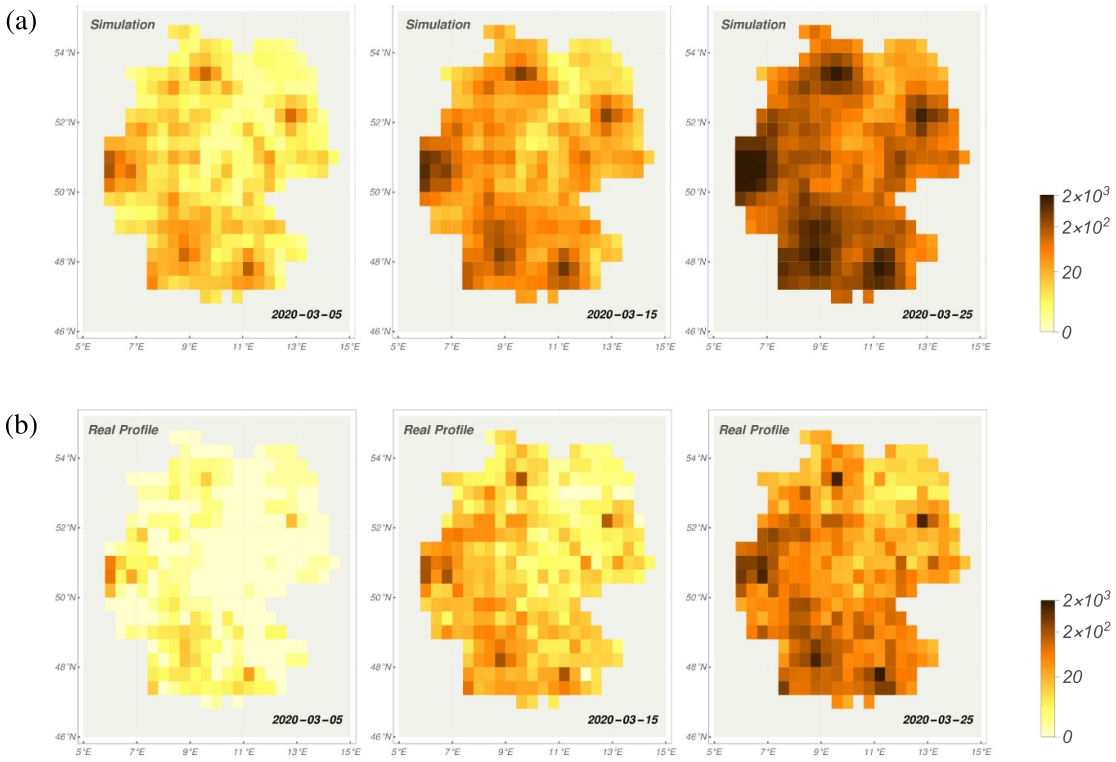

Figure 9. Comparison of simulated and RKI infection maps during March 2020 in Germany. The color bar is set in an uniform orange color scale, the darkest color is 2000 cases with a logarithmic re-scale.

assigned equally into each site. And if there is more than one county or county part on a site, the populations of the site is set as the sum of all these counties or county parts. Because the geographic distance is not important in our simulation, which focus on the population distributions, we set each site as a square when plotting the distribution in figure 8(right). In the current simulation, we confine all the population in Germany and thus set the out-of-Germany transfer possibility at the border as zero.

According to the initial infection cases distribution and the locations of main cities of Germany, we set the initial undiagnosed population distribution $U_0(m, n)$ to be proportional to the population distribution of the cities $N(m, n)$. In the simulations for March we choose the transfer rate $\varepsilon = 0.08$ as a constant, the population of transfer on site is proportional to the rate in each time-step. With the fitted parameters from above SUIR model, we generate the simulated evolution of the cumulative infection cases map in figure 9,

which demonstrates the high consistency with the RKI data map, named as real profiles. In which figure 9(a) shows that, as of the end of March, the cumulative infection cases are presented in dynamical simulation with fitting parameters at 2020-03-05, 2020-03-15 and 2020-03-25 respectively. Correspondingly, figure 9(b) shows the cumulative infection cases in real profiles.

Appendix B. SUIR model parameters explanation

In this part, we discuss the fitting details based on the mathematical simplification to the SUIR model. In the large susceptible population limit and ignoring the recovery and mortality cases for the uncertainty at early stage, there are only two equations are relevant,

$$\frac{dU}{dt} = (\lambda - \sigma p - \frac{\sigma(1-p)}{d})U \tag{5}$$

$$\frac{d(I+R)}{dt} = (\sigma p + \frac{\sigma(1-p)}{d})U, \tag{6}$$

where we redefine the $\sigma p + \frac{\sigma(1-p)}{d}$ as σ. The solutions are clearly derived as,

$$U(t) = U_0 Exp[(\lambda - \sigma)t] \tag{7}$$

$$I + R = I_0 + R_0 - \frac{\sigma U_0}{\lambda - \sigma} + \frac{\sigma U(t)}{\lambda - \sigma}. \tag{8}$$

The $I + R$ fitting is easy now with $I + R = a_* + b_* Exp[c_* t]$. This means we can not obtain all the independent parameters in the solution, but the following redefinition could be helpful,

$$a_* = I_0 + R_0 - \frac{\sigma U_0}{\lambda - \sigma} \tag{9}$$

$$b_* = \frac{\sigma U_0}{\lambda - \sigma} \tag{10}$$

$$c_* = \lambda - \sigma, \tag{11}$$

or equivalently only the following combinations could be obtained:

$$I_0 + R_0 = a_* + b_* \tag{12}$$

$$\sigma U_0 = b_* c_* \tag{13}$$

$$\lambda - \sigma = c_*. \tag{14}$$

As for the product σU_0, which means it is difficult to determine the σ and U_0 individually. It will introduce the inevitable uncertainty to the two parameters, while the confidence to the multiplicative variable could be reserved.

Appendix C. Conv2LSTM neural networks

The nuts-and-bolts of the neural network are: Convolution layers within TimeDistributed wrapper, from Convolution 2D (64, kernel size = 3 × 3, 'ReLU', 'same' padding) to Convolution 2D (64, kernel size = 3 × 3, strides = 2, 'ReLU', 'same' padding) to Convolution 2D (64, kernel size = 3 × 3, 'ReLU', 'same' padding) to Convolution 2D (64, kernel size = 3 × 3, 'ReLU', 'same' padding) and Flatten; for the LSTM part, 256 LSTM cells with 'sequence = False' and activation = 'tanh' are used; dense layers contains 256 neurons with 'ReLU' active function, before the output is the 'sigmoid' function to fit the normalized targets.

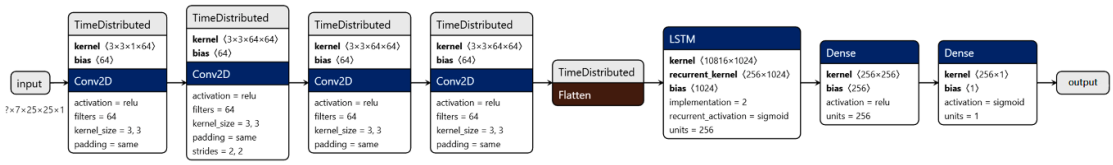

Figure 10. The Conv2LSTM neural networks.

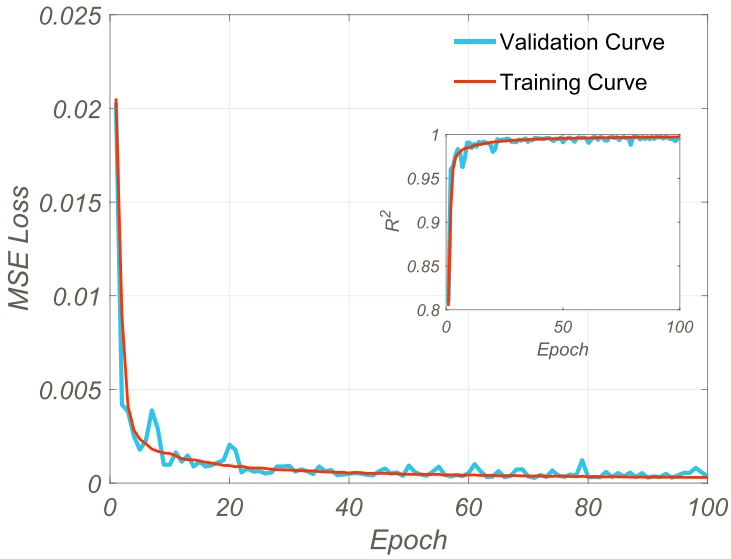

Figure 11. Learning curve of the Conv2LSTM neural networks. The red line is the learning curve, in which the mean-square error (MSE) loss is reducing with the epoch increasing. The blue line is located in validation data set, which shows the similar results in learning data set. The increasing of the R square with training is shown in subfigure, which indicates the correlation between the prediction from neural networks and truth tends to be close.

ORCID iDs

Lingxiao Wang https://orcid.org/0000-0003-3757-3403
Yin Jiang https://orcid.org/0000-0002-1916-6073
Kai Zhou https://orcid.org/0000-0001-9859-1758

References

[1] Maier B F and Brockmann D 2020 Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China *Science* **368** 742–6
[2] Walensky R P and del Rio C 1889 From mitigation to containment of the COVID-19 Pandemic: putting the SARS-CoV-2Genie back in the bottle *JAMA* **323** 2020
[3] Barbarossa M V, Fuhrmann J, Meinke J H, Krieg S, Varma H V, Castelletti N and Lippert T 2020 Modeling the spread of COVID-19 in Germany: early assessment and possible scenarios *PLoS One* **15** e0238559
[4] Barbarossa M V, Fuhrmann J, Heidecke J, Varma H V, Castelletti N, Meinke J H, Krieg S, Lippert T 2020 A first study on the impact of current and future control measures on the spread of COVID-19 in Germany (medRxiv)
[5] Estrada E 2020 COVID-19 and SARS-CoV-2. Modeling the present, looking at the future *Phys. Rep.* **869** 1–51
[6] Vespignani A *et al* 2020 Modelling COVID-19 *Nat. Rev. Phys.* **2** 279–81
[7] Wang L, Wang Y, Ye D and Liu Q 2020 Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence *Int. J. Antimicr. Agents* **55** 105948
[8] Kergassner A, Burkhardt C, Lippold D, Nistler S, Kergassner M, Steinmann P, Budday D, Budday S 2020 Meso-scale modeling of COVID-19 spatio-temporal outbreak dynamics in Germany (medRxiv)
[9] Linden M, Dehning J, Mohr S B, Mohring J, Meyer-Hermann M, Pigeot I, Schöbel A and Priesemann V 2020 The foreshadow of a second wave: an analysis of current COVID-19 fatalities in Germany (arXiv:2010.05850)
[10] Long Y, Chen Y and Li Y 2020 Multifractal scaling analyses of the spatial diffusion pattern of COVID-19 pandemic in Chinese mainland (arXiv:2010.02747)
[11] Chande A, Lee S, Harris M, Nguyen Q, Beckett S J, Hilley T, Andris C and Weitz J S 2020 Real-time, interactive website for US-county-level COVID-19 event risk assessment *Nat. Hum. Behav.* **4** 13
[12] Rader B *et al* 2020 Crowding and the shape of COVID-19 epidemics *Nat. Med.* **26** 1829–34
[13] Wang J, Lin X, Qilegeri, Liu Y, Lin H and Wu M 2020 Global dynamics of a SUIR model with predicting COVID-19 (arXiv:2004.12433)

[14] Zhou Y et al 2020 A spatiotemporal epidemiological prediction model to inform county-level COVID-19 risk in the United States *Harv. Data Sci. Rev.* (https://doi.org/10.1162/99608f92.79e1f45e)
[15] Gao Y et al 2020 Machine learning based early warning system enables accurate mortality risk prediction for COVID-19 *Nat. Commun.* **11** 5033
[16] Shen C, Chen A, Luo C, Zhang J, Feng B and Liao W 2020 Using reports of symptoms and diagnoses on social media to predict COVID-19 case counts in Mainland China: observational infoveillance study *J. Med. Internet Res.* **22** e19421
[17] Yeşilkanat C M 2020 Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm *Chaos Solitons Fractals* **140** 110210
[18] Zou D, Wang L, Xu P, Chen J, Zhang W, Gu Q 2020 Epidemic model guided machine learning for COVID-19 forecasts in the United States (medRxiv)
[19] White S H, del Rey A M and Sánchez G R 2007 Modeling epidemics using cellular automata *Appl. Math. Comput.* **186** 193–202
[20] Schneckenreither G, Popper N, Zauner G and Breitenecker F 2008 Modelling SIR-type epidemics by ODEs, PDEs, difference equations and cellular automata–a comparative study *Simul. Model. Practice Theory* **16** 1014–23
[21] Jun-feng L 2020 Assessment and prediction of COVID-19 based on SEIR model with undiscovered people *J. Univ. Electron. Sci. Technol. China* **49** 375–82
[22] Chen Y-C, Lu P-E, Chang C-S and Liu T-H 2020 A time-dependent SIR model for COVID-19 with Undetectable infected persons *IEEE Trans. Netw. Sci. Eng.* **7** 3279–94
[23] Ivorra B, Ferrández M R, Vela-Pérez M and Ramos A M 2020 Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China *Commun. Nonlinear Sci. Numer. Simul.* **88** 105303
[24] Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-k and Woo W-c 2015 ConvolutionalLSTM Network: A machine learning approach for precipitation nowcasting *Proc. 28th Int. Conf. Neural Information Processing Systems—1, NIPS'15* (Cambridge, MA: MIT Press) pp 802–10
[25] Ramos A A, Requerey I S and Vitas N 2017 DeepVel: deep learning for the estimation of horizontal velocities at the solar surface *A & A* **604** A11
[26] Wang L and Jiang Y 2019 Escape dynamics based on bounded rationality *Phys.* A **531** 121777
[27] Eckardt M, Kappner K, Wolf N 2020 Covid-19 across European regions: the role of border controls, type SSRN scholarly paper number ID 3688126 Social Science Research Network Rochester, NY
[28] Hernandez-Vargas E A and Velasco-Hernandez J X 2020 In-host mathematical modelling of COVID-19 in humans *Ann. Rev. Control* **50** 448–56
[29] Hartley G E, *et al* 2020 Rapid and lasting generation of B-cell memory toSARS-CoV-2 spike and nucleocapsid proteins in COVID-19 disease and convalescence (medRxiv)
[30] Moein S, Nickaeen N, Roointan A, Borhani N, Heidary Z, Javanmard S H, Ghaisari J and Gheisari Y 2021 In efficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan *Sci. Rep.* **11** 4725
[31] Hansen C H, Michlmayr D, Gubbels S M, Molbak K and Ethelberg S 2021 Assessment of protection against reinfection with SARS-CoV-2 among 4 million PCR-tested individuals in Denmark in 2020: a population-level observational study *Lancet* **397** 1204–12
[32] Pan Q, Gao T and He M 2020 Influence of isolation measures for patients with mild symptoms on the spread of COVID-19 *Chaos Solitons Fractals* **139** 110022
[33] Pan Q, Song S and He M 2021 The effect of quarantine measures for close contacts on the transmission of emerging infectious diseases with infectivity in incubation period *Phys.* A **574** 125993
[34] Gösgens M, Hendriks T, Boon M, Keuning S, Steenbakkers W, Heesterbeek H, van der Hofstad R and Litvak N 2020 Containment strategies after the first wave of COVID-19 using mobility data (arXiv:2010.14209)
[35] Vandenberg O, Martiny D, Rochas O, van Belkum A and Kozlakidis Z 2020 Considerations for diagnostic COVID-19 tests *Nat. Rev. Microbiol.* **19** 171–83
[36] Li Y, Yao L, Li J, Chen L, Song Y, Cai Z and Yang C 2020 Stability issues of RT-PCR testing of SARS-CoV-2 for hospitalized patients clinically diagnosed with COVID-19 *J. Med Virol.* **92** 903–8
[37] 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
[38] Tian H *et al* 2020 An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China *Science* **368** 638–42
[39] Bavel J J V *et al* 2020 Using social and behavioural science to support COVID-19 pandemic response *Nat. Hum. Behav.* **4** 460–71
[40] Hsiang S *et al* 2020 The effect of large-scale anti-contagion policies on the COVID-19 pandemic *Nature* **584** 262–7
[41] Wong G N, Weiner Z J, Tkachenko A V, Elbanna A, Maslov S and Goldenfeld N 2020 Modeling COVID-19 dynamics in Illinois under nonpharmaceutical interventions *Phys. Rev.* X **10** 041033
[42] Chang S, Pierson E, Koh P W, Gerardin J, Redbird B, Grusky D and Leskovec J 2020 Mobility network models of COVID-19 explain inequities and inform reopening *Nature* **589** 82–7
[43] Stafford N 2020 Covid-19: why Germany's case fatality rate seems so low *BMJ* **369** m1395
[44] Annas S, Isbar Pratama M, Rifandi M, Sanusi W and Side S 2020 Stability analysis and numerical simulation ofSEIR model for pandemic COVID-19 spread in Indonesia *Chaos Solitons Fractals* **139** 110072
[45] Topirceanu A, Udrescu M and Marculescu R 2020 Centralized and decentralized isolation strategies and their impact on the COVID-19 pandemic dynamics (arXiv:2004.04222)
[46] Valba O, Avetisov V, Gorsky A and Nechaev S 2020 Self-isolation or borders closing: what prevents the spread of the epidemic better? *Phys. Rev.* E **102** 010401
[47] Ye Y, Zhang Q, Ruan Z, Cao Z, Xuan Q and Zeng D D 20220 Modeling the heterogeneous disease-behavior-information dynamics during epidemics (arXiv:2005.07012)
[48] Charoenwong B, Kwan A and Pursiainen V 2020 Social connections with COVID-19-affected areas increase compliance with mobility restrictions *Sci. Adv.* **6** eabc3054