

Article

COVID-19 Spatial Diffusion: A Markovian Agent-Based Model

Marco Gribaudo ^{1,*}, Mauro Iacono ² and Daniele Manini ³

¹ Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milano, Italy

² Dipartimento Matematica e Fisica, Università degli Studi della Campania “Luigi Vanvitelli”, 81100 Caserta, Italy; mauro.iacono@unicampania.it

³ Dipartimento di Informatica, Università degli Studi di Torino, 10149 Torino, Italy; manini@di.unito.it

* Correspondence: marco.gribaudo@polimi.it

Abstract: We applied a flexible modeling technique capable of representing dynamics of large populations interacting in space and time, namely Markovian Agents, to study the evolution of COVID-19 in Italy. Our purpose was to show that this modeling approach, that is based on mean field analysis models, provides good performances in describing the diffusion of phenomena, like COVID-19. The paper describes the application of this modeling approach to the Italian scenario and results are validated against real data available about the Italian official documentation of the diffusion of COVID-19. The model of each agent is organized similarly to what largely established in literature in the Susceptible-Infected-Recovered (SIR) family of approaches. Results match the main events taken by the Italian government and their effects.

Keywords: COVID-19; spatial diffusion model; Markovian Agents



Citation: Gribaudo, M.; Iacono, M.; Manini, D. COVID-19 Spatial Diffusion: A Markovian Agent-Based Model. *Mathematics* **2021**, *9*, 485. <https://doi.org/10.3390/math9050485>

Academic Editor: Tuan Phung-Duc

Received: 28 December 2020

Accepted: 20 February 2021

Published: 26 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

News outlets report daily numbers about COVID-19 spread, evolution, and consequences. In the very first months, the media was hectically running after data published by authorities, that often were confused, or related to measures that were taken with different criteria, or reported with different delays and missing proper timestamping. The crisis found the international community dis-aligned and without unique modus operandi for data collection, metadata organization, validation, or information extraction: moreover, the behavior and the reactions of governments has been quite different and without coordination. In some cases, in which local authorities are directly in charge of decisions about COVID-19 management, the situation became fragmented and odd phenomena emerged, causing confusion, misleading information, chaotic management of the situation, and delays in decisions and in enacting solution strategies. Politicians turned to scientists for advice, but there was no unanimous response, and a certain spectacularization of scientific communication caused doubts and bewilderment in public opinion. Lockdowns mitigated the consequences of the initial circulation of the virus but, in turn, had severe consequences on the economy of many countries, endangering many enterprises and putting small businesses and shops that depend on normal circulation of people in serious trouble because they cannot survive the costs of prevention and the change of habits. For example, the development of the events in the case of Italy has been briefly, but richly, resumed in Reference [1], in which a detailed chronological description of the events is reported, together with the main implications and a model of COVID-19 epidemic and consequent interventions in Italy founded on an extended Susceptible-Infected-Recovered (SIR) approach.

These months produced an enormous volume of scientific literature about whatever is related to COVID-19, from clinical reports to policy-oriented prescriptions in social sciences: what seems to emerge as a fact is that COVID-19 is still here to stay and that quantitative modeling of COVID-19 is crucial to cope with the management of the crisis and

to support resource allocation at every level and planning of reactions. Providing predictive models to governments and international organization is of paramount importance to control damages and delay the second wave, that is documented in many countries, until enough clinical knowledge will be assessed and consolidated to produce optimal protocols for treatment, proper clinical setups for patient management, and effective medicines, hopefully including preventative means.

Modeling the spread of the virus to provide predictive tools should consider many aspects. The first aspect to be considered is the evolution in time of the phenomenon: anyway, it does not capture its dynamics, which are not completely described by levels that change in time and thresholds, but need to take into account state changes, that document relations between the conditions of a same individual of the population in different time frames. The analyses reported in Reference [2,3] provide good insights and data for COVID-19. Read et al. [2] used simulation to evaluate the virus impact on Wuhan and other cities of China. Anastassopoulou et al. [3] estimate case fatality and recovery ratios using data on Wuhan outbreak and a variant of an SIR model.

Scientific literature largely considered the problem of mathematical models for the description of infectious diseases and epidemiological dynamics. The topic is object of comprehensive analyses and discussions that we recommend to interested readers as general introduction references. For example, Diekmann and Heesterbeek [4] propose translating biological assumptions into mathematics, focusing on a perspective of mathematical analysis aided by interpretation and about obtaining insight into epidemic phenomena when translating mathematical results back into population biology. Hethcote [5] presents endemic models with either continuous age or age groups parameters and also considers various diseases. Brauer and Castillo-Chavez [6] consider simple and analyzable models and unsolvable models to address the population explosion under different conditions. Finally, Cordero et al. [7] describe a signal transduction pathway involved in a biological process by a detailed representation which results in mathematical models with a lower computational and structural complexity.

As also evidenced by the previous citations, a popular family of models is that of SIR models [8], which are based on the classification of individuals of the population into categories (the name derives from a classification into three groups: Susceptible, Infected, and Recovered) that can be considered as states between which individuals move according to given probabilities that are inferred by existing statistical data, or more complex mechanisms, including contact networks, which are, anyway, a problem worth of investigation that literature dealt with, as the availability of contact data is generally rare and not fully reliable. In the following, we list some works focused on contact data: Machens et al. [9] show that a contact matrix that only contains average contact durations between role classes fails to reproduce the size of the epidemic obtained using the high-resolution contact data and also fails to identify the most at-risk classes; Sapiezynski et al. [10] and Huang et al. [11] provide multiple types of communication and contact networks, respectively; McCaw et al. [12] study different mathematical models of infection that are dependent on the assumed mixing patterns of the population; Salathe et al. [13] analyze infectious diseases in an American college spread by close proximity interactions. Moreover, the Italian government promoted an app for smartphones, namely Immuni [14], to collect anonymous information about contacts based on Bluetooth proximity, with some minor results, and there is a request for properly organizing the existing data about the pandemics and its treatment in the spirit of the OpenData initiative, to disclose information for community-based research efforts, including network-based modeling. SIR is a family of models that is widely explored and used in literature, and is the base for many variants, such as SIS, after Susceptible, Infected, and Susceptible, SEIR, that adds a Exposed stage, or, for example, the recent SIDARTHE [1] (Susceptible, Infected, Diagnosed, Ailing, Recognized, Threatened, Healed, and Extinct) and SECIHURD [15] (Susceptible, Exposed, Carrier, Infected, Hospitalized, in intensive care Unit, Dead, and Recovered) specifically developed for COVID-19. The model described in Reference [1] deals with data about

Italy, uses a larger number of stages for the evolution of COVID-19 to provide a closer representation of the phenomenon, and is a case that represents how researchers evolved the basic idea into different custom variants, while the model described by Großmann et al. [15] shows how an SIR-like model can be translated into a meanfield model for solution. Another recent example of application of SIR models, using data from China, is provided by Chen et al. [16], in which interested readers can find a good step-by-step description of the approach, that we found useful to get insights into this modeling technique, while Reference [8] provides a very good introduction to modeling of epidemic processes in complex networks in general.

An important feature that models can leverage to increase their modeling power and significance with respect to the transmission dynamics of the virus is their accounting for spatial features. In fact, considering all individuals as equally subject to contagion with no respect for their presence in places with a higher or lower concentration of infection vectors, that is, in a densely populated environment rather than in a convenient and prudential isolation condition, going beyond contact networks, increases the realism. The movements of infected subjects on the considered territory and their presence on board of public transportation, or the speed of their movement across the territory, is a relevant aspect to be considered: the spread of infection in Italy during the first phase was largely asymmetric across different regions because contagion was concentrated in northern areas and traveling was controlled, but a sudden and early announcement of regional lockdown caused southern people working in northern regions to assault the last southbound trains, paving the way for a rise of infection cases in the south, as well. Please note that regions in Italy have local autonomy and administrative powers, that include the possibility of limitations of free movements between regions and even smaller administrative subdivisions: in the most critical phase of the lockdown, people was forbidden to exit the borders of their municipalities without a certificated need.

Spatial models may account for these factors. Spatial models consider locality when describing the probability of virus transmission, and describe subjects as agents that may move from one place to another, may get in contact in different places with different agents density, or may be subject to a different exposition to the virus in places that represent situations in which higher or lower risk mitigation actions are taken. Considering agents evolution in time and space and their actual contacts (that is, coexistence in the same place in the same time) may provide a more realistic evaluation of specific behaviors proper of different agents classes (e.g., different categories of workers, different perception of the need for respecting the rules, different requirements in visiting some specific places of the model, different frequency of presence or persistence in specific places). In this paper, we adopt a modeling technique that has spatial features, namely Markovian Agents [17], that well matches SIR-like models: first of all, it is based on agents that stochastically evolve between states coherently with the SIR logic, in which compartments can be associated to the agent model states. Furthermore, individual behaviors, which are affected by global interactions with other individuals and time-related location-dependent events, are well suited to be defined with a parametric grain of the interactions in the space (an example of such an application of the modeling technique is provided by Gribaudo et al. [18]). To highlight the contribution of this work to the literature, please refer to the works presented in Reference [1,15]. The former take into account the Italian contagion as we did in our work, while the latter develop an approach starting from the mean field analysis that we have adopted as a solution technique to calculate the results we presented. A relevant difference is that our study is able to consider the spatial location of the population and, in particular, the impact of the infection among the Italian regions, beyond the overall national evolution of the pandemic presented in Reference [1] that our model can also represent. Furthermore, the goal of Großmann et al. [15] is to point out that stochastic models offer a more reliable epidemic prediction tool than deterministic ones. We exploit a formalism, namely the Markovian agent-based models, and the provided results show a good fit to the real data provided by the government.

The aim of this work was to show how these formalisms, as agent-based Markov models, are suitable for describing and evaluating systems characterised by enormous quantities of entities, such as countries where the population is affected by a pandemic. In this case, the evolution of the COVID-19 epidemic was studied with the support of real data to understand the applicability of such approaches.

2. Materials and Methods

To prove the increase potential of spatial-based modeling in cases, like the study of the evolution of COVID-19, we start from the extended SIR model presented in Reference [15], and we tailor it to match the measurements available in Reference [19]. In particular, we consider the following states for each individual:

- **S**—Susceptible: the individual has never been in contact with the virus and can still be infected.
- **I**—Infected: the individual has been infected by COVID-19. However, she is still *asymptomatic* and remains *undetected* showing no evidence of the infection. An individual in this state can infect other people and can either be recognized as infected, recover autonomously, or die without knowing that his death was caused by the virus.
- **Q**—Home treatment: the individual has been *detected* as infected, but he/she is still in a condition that can be managed from his/her own home.
- **H**—Hospitalized: the health conditions become more severe, requiring hospitalization.
- **C**—Intensive Care: the health conditions are critical, requiring intensive care treatment.
- **R** and **R'**—Recovered: the individual has been recognized as recovered, and he/she is no longer in danger or contagious. Measurements only account for recovered individuals that have been detected. To match available data, we split the recovered state into two: the first (**R**) accounts for the recovery of individual who were discovered, while the second (**R'**) considers the recovered individuals which remained undetected.
- **D** and **D'**—Dead: the individual who died due to the virus. In addition, in this case, the state is split into two, to consider both the ones that have been detected (state **D**) and the ones that died before their infection was discovered (state **D'**).

The main purpose of this model is to match data that we can find in the public available data set, which report the evolution of the number of people in states **Q**, **H**, **C**, **R**, and **D** for each day of the evolution of the spread of the disease.

The proposed model can be directly mapped on the Markovian Agent shown in Figure 1. Each state of the SIR model is directly mapped into an agent state. Agents are further divided into classes, all with the same structure shown in Figure 1 but with different parameters: in particular, we used one agent class for each Italian region. Note that transitions among states are determined by specific actions that are indicated in the diagram by arcs with different colors; the text in the figure shows the association action/arc color.

The transition rate from state *S* to state *I* is proportional to the total number of infected people the agent might come by, which includes both his local neighbors, plus agents in other regions that might come in contact with him. In particular, let us focus on a given region *c*. We call $q_{SI}^{(c)}(t)$ the transition rate from state *S* to state *I* at time *t*, and we defined it as:

$$q_{SI}^{(c)}(t) = \beta(t) \cdot \mathcal{I}^{(c)}(t), \quad (1)$$

where parameter $\beta(t)$ is a location-independent, but time-dependent, parameter that accounts for the infection rate. We assume it to be independent from the location, since it is connected to the specific infectiveness of the considered virus, which we assume being constant over all the considered territory. Instead, we make it time-dependent, to include the reduction of exposure due to countermeasures applied by social distancing and hygiene policies. $\mathcal{I}^{(c)}(t)$ represents the total number of individual which might become in contact with the considered agent in location *c* at time *t*: we will return on its definition later. Let

us define the following rates: η the rate at which an infected person gets recognized, $\zeta(t)$ the death rate, and $\gamma(t)$ the rate at which an individual recovers. Moreover, a recognized agent can be cared for at home with probability $p_Q(t)$, hospitalized with probability $p_H(t)$, and requiring intensive care with probability $p_C(t)$. The transition rates out of the infected state **I**, leading to states **R'** (recovered without having been discovered), **D'** (death without having been discovered), **Q** (home treatment), **H** (Hospitalized), and **C** (requiring intensive care), can be, respectively, defined as follows:

$$q_{IR'}^{(c)}(t) = \gamma(t), \tag{2}$$

$$q_{ID'}^{(c)}(t) = \zeta(t), \tag{3}$$

$$q_{IQ}^{(c)}(t) = p_Q(t) \cdot \eta, \tag{4}$$

$$q_{IH}^{(c)}(t) = p_H(t) \cdot \eta, \tag{5}$$

$$q_{IC}^{(c)}(t) = p_C(t) \cdot \eta. \tag{6}$$

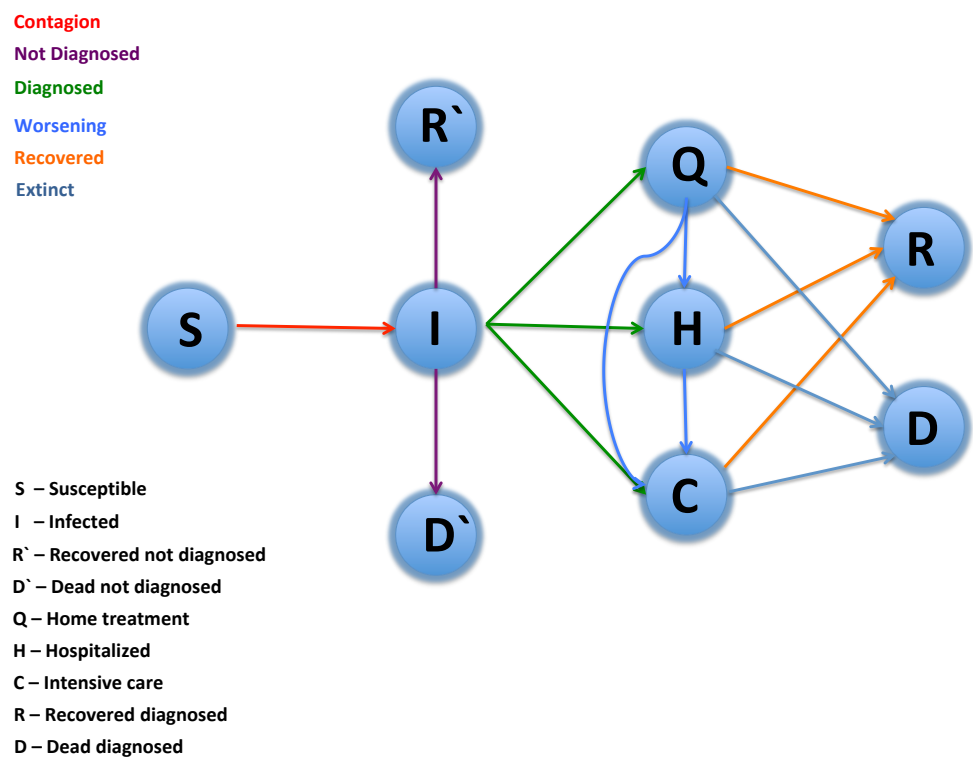


Figure 1. COVID-19 model.

Rates $\gamma(t)$ and $\zeta(t)$ and probabilities $p_Q(t)$, $p_H(t)$, and $p_C(t)$ have been considered time-dependent to model the improvements in the treatment that reduces the recovery time and chance of death as time passes. Rates $\gamma(t)$ and $\zeta(t)$ are also used to describe the recovery and death from states different from *I*, i.e.,

$$q_{QR}^{(c)}(t) = q_{HR}^{(c)}(t) = q_{CR}^{(c)}(t) = \gamma(t), \tag{7}$$

$$q_{QD}^{(c)}(t) = q_{HD}^{(c)}(t) = q_{CD}^{(c)}(t) = \zeta(t). \tag{8}$$

Worsening of the disease is modeled with time dependent parameters $\tau_H(t)$ and $\tau_C(t)$, by defining jump rates from state **Q** (home treatment) to either **H** or **C** (hospitalized or intensive care), and from **H** to **C**, as follows:

$$q_{QH}^{(c)}(t) = \tau_H(t), \tag{9}$$

$$q_{QC}^{(c)}(t) = \tau_C(t), \tag{10}$$

$$q_{HC}^{(c)}(t) = \tau_C(t). \tag{11}$$

The key variable of this model is $\mathcal{I}^{(c)}(t)$: the total number of individual which might become in contact with the considered agent. If we call $n_S^{(c)}(t)$ the number of agents in state S of region c at time t , we used:

$$\mathcal{I}^{(c)}(t) = \sum_{regions:d} w_{cd}(t) \left[n_I^{(d)}(t) + \alpha_{QHC} \left(n_Q^{(d)}(t) + n_H^{(d)}(t) + n_C^{(d)}(t) \right) \right]. \tag{12}$$

Equation (12) accounts for the fact that an agent can be infected if it comes in contact with an agent in states I, Q, H , or C . Coefficient α_{QHC} is used to give a different weight to agents that are known to be infected since they are either in states Q, H , or C . Factor $\mathcal{I}^{(c)}(t)$ accounts also for agents in other regions, according to a weight $w_{cd}(t)$, which depends on the considered region c and each neighbor d . In particular, $w_{cc}(t) = 1$ accounts for the agents in the same area, while $w_{cd}(t) \rightarrow 0$ as neighbor regions become more distant. This weight $w_{cd}(t)$ is also time dependent to model the effect of confinement and lock-down policies.

2.1. Parameter Definition

We have focused on eight Italian regions and determined the model parameters with either a fitting procedure on the data available in the local civil protection service or from public national databases. Note that, by fitting the parameters with data sets of other countries, the considered model may be used for different geographical areas.

In the following, we will present the obtained values, and the fitting procedure will be explained at the end of this section. The considered regions, namely Piemonte (denoted with PI), Valle d’Aosta (VA), Lombardia (LO), Trentino Alto Adige (TA), Veneto (VE), Friuli (FR), Liguria (LI), and Emilia Romagna (ER), are the ones who played a very important role in the beginning of the virus spreading process. We started the analysis on 24 February 2020 on day 0, and then we modulated the parameters of the model according to the different time phases foreseen by the Italian government, reported in Table 1.

Table 1. Phases from day_0 February the 24th 2020.

Phase Start	Italian Phase Name	Phase Rule	Infection Rate β
day_0	Giorno zero	Day zero	3.083×10^{-8}
$day_0 + 7$	Zona rossa	Red Zone	2.316×10^{-8}
$day_0 + 14$	Io resto a casa	I stay at home	1.022×10^{-8}
$day_0 + 29$	Chiudi Italia	Italian lockdown	0.666×10^{-8}
$day_0 + 71$	Fase 2	Phase 2	1.763×10^{-8}
$day_0 + 85$	Nuove riaperture	Additional re-opening	0.246×10^{-8}
$day_0 + 101$	Mobilità tra regioni	Mobility between regions	0.25×10^{-8}

As introduced in Equation (1), infection rate parameter β is time dependent and, in particular, its values vary according to the considered time phases. Table 1 illustrates the temporal values obtained by the fitting procedure described below.

Figure 2 shows the estimated initial number of contagions for each region, including the no-diagnosed (yellow bar labeled with the value), and also the diagnosed divided in home treated (blue), hospitalized (orange), and in intensive care (gray). The diagnosed are taken from Reference [19], while the no-diagnosed are computed by multiplying the number of diagnosed with a coefficient ϵ_0 determined with fitting. In particular, we have

set it to 5.32. Note that this is not too different to the value 6, which was supposed in most of the literature, e.g., Reference [20]. The initial number of susceptible has been derived subtracting the number of infected individuals accounted in the other states from the number of inhabitants of the region, reported in official documents and shown in Table 2.

Table 2. Regions population. Reproduced from www.istat.it.

Region	Population
PI	4,356,406
VA	125,666
LO	10,060,574
TA	1,072,276
VE	4,905,854
FR	1,215,220
LI	1,550,640
ER	4,459,477

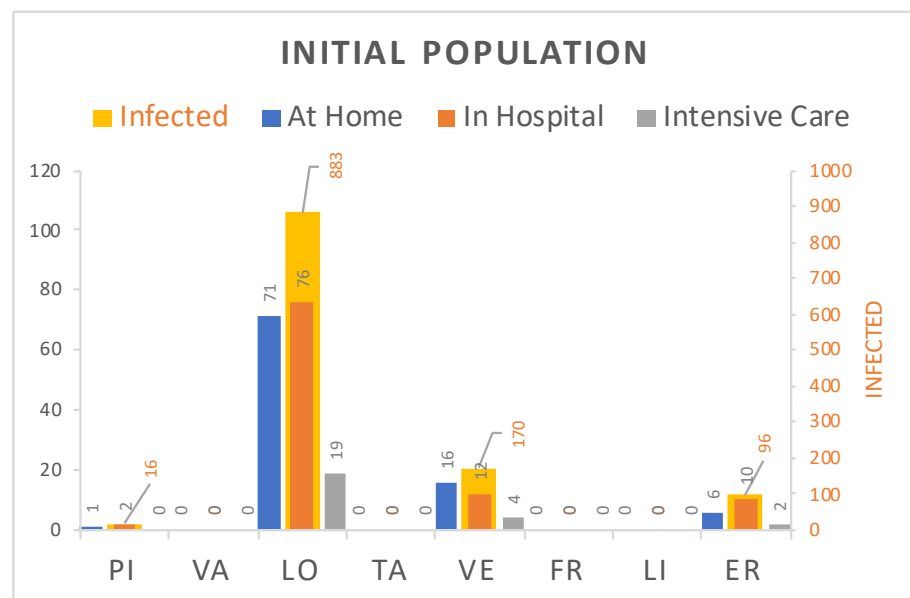


Figure 2. Initial agent distribution.

As mentioned above, contagion in a given region c also counts for agents in neighboring regions d according to a weight of $w_{cd}(t)$. We have obtained these weights as the product of two factors:

$$w_{cd}(t) = \omega_{cd} \cdot l(t), \tag{13}$$

where ω_{cd} considers the geographical distance, and, in some cases, we have set them by fitting to take into account particular correlations, as for LO, which in fact strongly determined the spread of the virus in the neighboring regions. The resulting coefficients w_{cd} are reported in Table 3. The effect of lockdown is modeled by $l(t)$ and is set to 1 in time phases without lockdown, and to 0.1 otherwise.

Table 3. Parameters w_{cd} that consider the dependence of the contagion between regions.

	PI	VA	LO	TA	VE	FR	LI	ER
PI	1	0.9971	0.6131	0	0	0	0.6431	0
VA	0.9971	1	0.8993	0	0	0	0.3188	0
LO	0.5505	0.3871	1	0.3385	0	0	0.8802	0.2445
TA	0	0	0.6594	1	0.6750	0.1834	0.1756	0.5498
VE	0	0	0.1240	0.6750	1	0.6218	0	0.6816
FR	0	0	0	0.1834	1.4800	1	0	0.1839
LI	0.6431	0.3188	0.5400	0.1756	0	0	1	0.2322
ER	0	0	0.3663	0.5498	0.6816	0.1839	0.2322	1

By using the fitting technique which will be introduced in the next section, we also define the following parameters. The rate η at which an infected person gets recognized has been set to 0.05. This speed is modulated by two sets of probabilities depending on the date: values $p_Q = 0.58$, $p_H = 0.35$, and $p_C = 0.07$ are used to determine if diagnosed contagions become home treated, hospitalized, or in intensive care before the lockdown, and $p_Q = 0.83$, $p_H = 0.16$, and $p_C = 0.01$ are used for the same cases later.

Rates that account for the disease worsening, i.e., $\tau_H(t)$ and $\tau_C(t)$, have been, respectively, set to 0.025 and 10^{-4} before the lockdown, and to $1/900$ and 10^{-5} later. The first parameter, $\tau_H(t)$, defines the speed at which home treated become hospitalized, and the second one, $\tau_C(t)$, accounts for both home treated and hospitalized going in intensive care.

The recovery rate γ and death rate ζ depend on time to consider the changes in the logistical situation of the hospitals and the improvement in medical care that occurred during the pandemic. γ has been set to 0.0762 from day zero to the beginning of the Italian Lockdown, then to 0.0198 until Phase Two started, and, finally, to 0.03. Death rate ζ has been set to 0.0035 until the beginning of the Italian Lockdown, and to 0.0015 until the end of analysis.

2.2. Parameter Fitting

To show the potentiality of the spatial agent-based modeling formalism, we fit the parameters with the measurements available from the Italian Civil Protection. We used a custom genetic algorithm implemented in MathWork’s MATLAB R2020a Update 3 (May 27 2020), which focuses on a single time-varying parameter at a time. We used the time instants defined in Table 1 and considered each interval t_i (for a total of $|\{t_i\}|$ intervals). For each parameter $p_x(t)$, we identified which state y and region c were more influenced by it, as well as minimized the distance between real measures $m_y^c(t)$ and model prediction $n_y^c(t)$ at the end of considered interval $t_i + 1$. Algorithm 1 shows a simplified version of the procedure used.

Algorithm 1 Fit parameter $p_x(t)$ against real measures $m_y^{(c)}(t)$ using reference times t_i

```

Set current error  $\epsilon = \infty$ 
for  $k=1\dots\text{MaxIterations}$  do
  for  $i = 1\dots(|\{t_i\}| - 1)$  do
    Set a random perturbation  $p'_x(t_i)$  of  $p_x(t_i)$  :
       $p'_x(t_i) = \min(\max(p_x(t_i) \cdot \text{random}[v_l, v_u], p_{\min}), p_{\max})$ 
    Solve the model with  $p'_x(t_i)$ 
     $\epsilon' = |m_y^c(t_{i+1}) - n_y^c(t_{i+1})|$ 
    if  $\epsilon' < \epsilon$  then
       $\epsilon = \epsilon'$ ,  $p_x(t_i) = p'_x(t_i)$ 
    end if
  end for
end for

```

We run several executions of the algorithm to perform a random restart, to avoid that the procedure may converge to a local minimum. In particular, we restarted the algorithm 10 times: in each execution, setting $MaxIterations = 300$ was sufficient to find a good set of values. The random perturbation was performed by multiplying the previous value of the parameter with a uniformly distributed value in a range $[v_l, v_u]$ with $v_l < 1$ and $v_u > 1$, and a constraint to keep the resulting value in an admissible range $[p_{min}, p_{max}]$ to avoid finding results that are too far from the previous versions.

3. Results

Using a Markovian Agent-based description of the virus evolution, we have been able to obtain results that are very close to the ones reported by the Italian authorities. The main advantage of the results obtained is that they are able to consider each region in a different way, as shown in Figure 3, where the evolution of the contagion is reported. In particular, the picture compares the results of the simulation obtained solving our model, with the real data for what concerns the total number of infected people, the number of detected cases, and the number of deaths for each day in the considered period. As it will be described later, our model accounts for a constant geographical influence, determined mainly by the ongoing activities between the different areas, and modulated by political restrictions, such as the lockdown period. Vertical lines in the plots reported in Figures 3 and 5 represent the days when the government announced new national strategies (see Table 1). Parameters regarding the capacity of the infection to spread and to be detected, and also about the recovery and worsening of the cases, are instead modulated with global (region independent) time-dependent values, to reflect both the effects of different social distancing policies and the improved capability of doctors to diagnose and cure the disease.

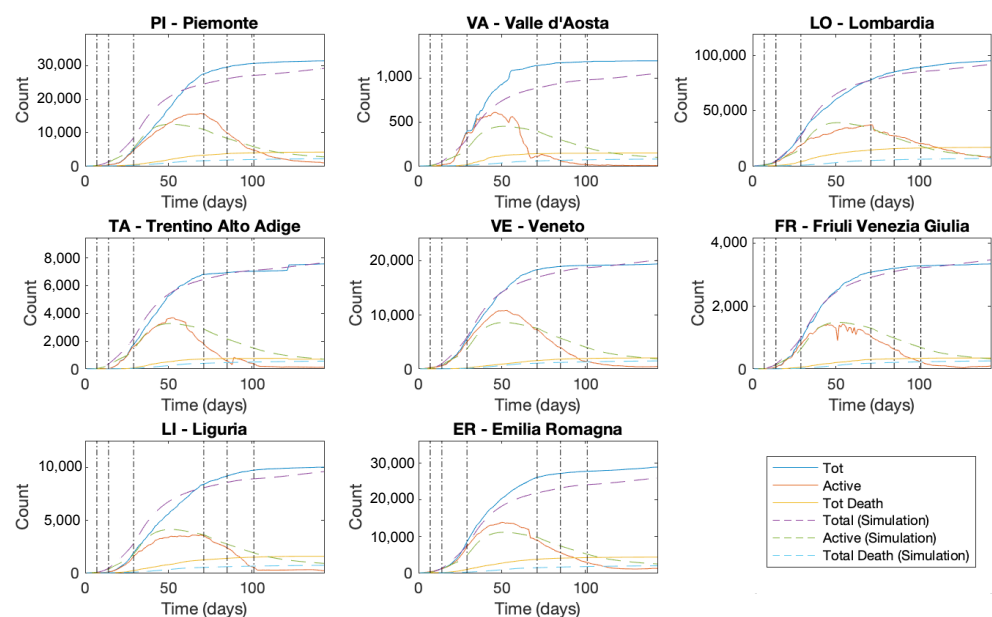


Figure 3. Real plots versus model results for each region.

Figure 4 shows the average relative error among regions of the simulation approach against the real life data. It can be noted that, at the beginning, matching is optimized against the active cases, while, at the end, the total cases matching is better.

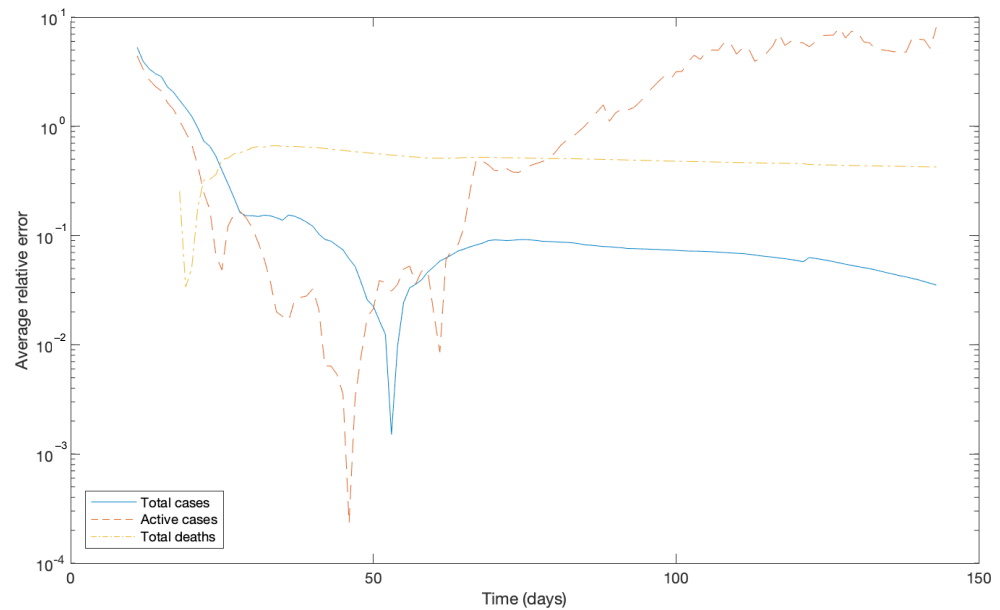


Figure 4. Average relative error of the simulation approach against the real life data in log scale.

Results can be aggregated to consider the total evolution over the whole considered territory, as shown in Figure 5. In this case, we also present the distribution of infected cases among home treatment, hospitalization, and intensive care. As it can be seen, aggregated results remain very close to the real ones also for the latter measures.

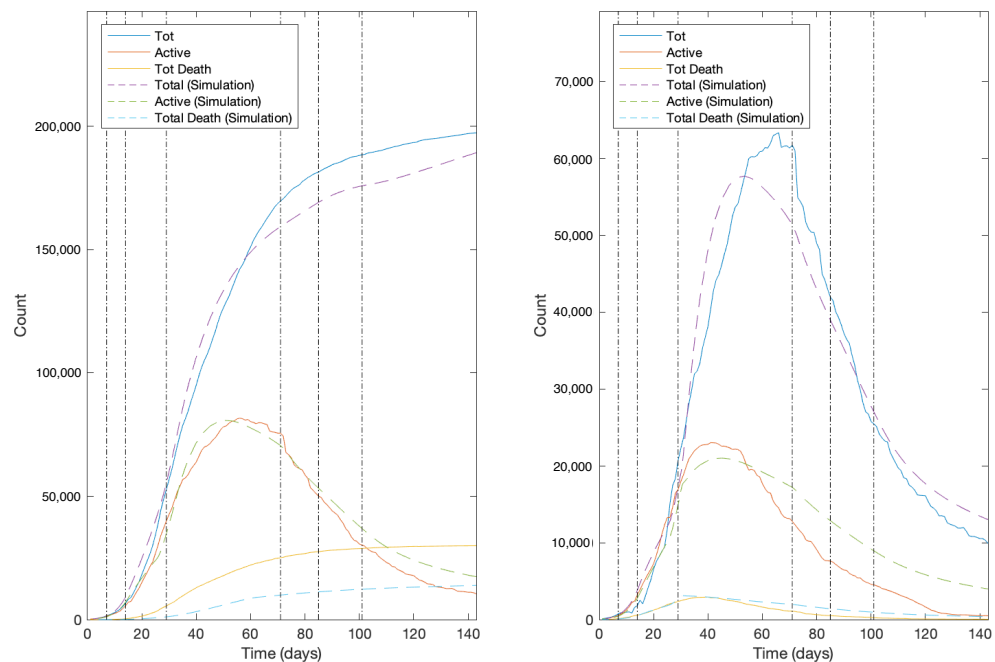


Figure 5. Real plots versus model results for the whole Italy, and analysis of at home, at hospital, and intensive care patients.

4. Discussion

The model we developed simulates the agents evolution in time and space, and it allowed us to represent a realistic scenario where the probability of a virus transmission in different places can be considered. The proposed technique is based on mean field analysis that provides a more precise evaluation for the dynamics of a large population of agents as in the case of a pandemic. We computed the spread of the COVID-19 for eight Italian regions taking into account the contagion among all their populations (see Figure 3);

however, the model can be easily extended to consider all twenty-one regions. Moreover, the contagion in the whole country was also simulated (see Figure 5). All results present a good accuracy when compared with the available information, as it can be seen in Figure 4, where the average relative error of the model results against the real life data is plotted.

This approach could lead to new experiments in which the effect of vaccine inoculation can be analyzed both locally and nationwide, and it can also be used to observe the impact of outbreaks.

Simulations were conducted solving a Markovian Agent model, which represents a system as a collection of *Markovian Agents* (MAs) spread over a geographical space \mathcal{C} which, in this case, corresponds to the set of Italian regions with $K = |\mathcal{C}|$. Each MA is described by a Continuous Time Markov Chain, in which an infinitesimal generator describes the transition rates among N different states.

Agents are analyzed by studying the evolution of their *counting process* and exploiting mean field approximation [21,22]. In particular, the average number agents in state s for a region $c \in \mathcal{C}$ at time t is denoted as $n_s^{(c)}(t)$. To simplify the notation, let us collect the terms into row vectors $\mathbf{n}^{(c)}(t) = \langle n_s^{(c)}(t) \rangle$ that represent the state distribution of a MA belonging to region c at time t . The dynamic evolution of the average number of agents can be computed solving the system of ordinary differential equations in (14):

$$\frac{d\mathbf{n}^{(c)}(t)}{dt} = \mathbf{Q}^{(c)}(t) \cdot \mathbf{n}^{(c)}(t). \quad (14)$$

Here, the elements of matrix $\mathbf{Q}^{(c)}(t)$ may depend on the count of the population of region c , and of other neighbor regions, as well as from time t . In particular, $\mathbf{Q}^{(c)}(t) = [q_{ij}^{(c)}(t)]$ represents the transition rate from state i to state j in a region c at time t .

The considered scenario leads to a small system of $N \cdot K$ well conditioned differential equations, where N is the number of states and K the number of regions. We performed its solution in MATLAB using `ode45(...)` solver, and it required only few milliseconds.

5. Conclusions

In this preliminary study, we showed that Markovian Agent Models, thanks to their ability to include both spatial issues and individual behaviors that are affected by global interactions, are well suited to model the distributed evolution of COVID-19. The intrinsic characteristics of this approach allow the researchers to obtain a more precise representation of systems with a large population of entities as in the case of a pandemic. We have focused on eight Italian regions and determined the model parameters with either a fitting procedure on the data available in the local civil protection service or from data recorded in public national databases. The obtained results are very close to the ones reported by the Italian authorities.

Future works will aim at an even finer grained scale, considering the evolution of provinces instead of regions and capturing a more detailed spatial evolution. We would also like to exploit the model to check the effectiveness of new measures in case of successive infection waves. For example, we aim to evaluate the effects of vaccine administration on a regional and national scale, as well as to model the impact of outbreaks according to their spatial position and dimension. In particular, using the same formalism we could include supply shortage, vaccine distribution, balancing between regions, and the evaluation of inoculation intervals.

Author Contributions: Conceptualization, M.G., M.I. and D.M.; methodology, M.G., M.I. and D.M.; software, M.G., M.I. and D.M.; validation, M.G., M.I. and D.M.; formal analysis, M.G., M.I. and D.M.; investigation, M.G., M.I. and D.M.; resources, M.G., M.I. and D.M.; data curation, M.G., M.I. and D.M.; writing—original draft preparation, M.G., M.I. and D.M.; writing—review and editing, M.G., M.I. and D.M.; visualization, M.G., M.I. and D.M.; supervision, M.G., M.I. and D.M.; project administration, M.G., M.I. and D.M.; funding acquisition, M.G., M.I. and D.M.. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in www.istat.it, and in reference number [19].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Giordano, G.; Blanchini, F.; Bruno, R.; Colaneri, P.; Di Filippo, A.; Di Matteo, A.; Colaneri, M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* **2020**, *26*, 855–860.
2. Read, J.; Bridgen, J.R.; Cummings, D.A.T.; Ho, A.; Jewell, C.P. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv* **2020**, doi:10.1101/2020.01.23.20018549.
3. Anastassopoulou, C.; Russo, L.; Tsakris, A.; Siettos, C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS ONE* **2020**, *15*, e0230405.
4. Diekmann, O.; Heesterbeek, J.A.P. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*; Wiley: Chichester, UK, 2000.
5. Hethcote, H.W. The mathematics of infectious diseases. *SIAM Rev.* **2000**, *42*, 599–653.
6. Brauer, F.; Castillo-Chavez, C. *Mathematical Models in Population Biology and Epidemiology*, 2nd ed.; Springer: New York, NY, USA, 2012.
7. Cordero, F.; Horváth, A.; Manini, D.; Napione, L.; De Pierro, M.; Pavan, S.; Picco, A.; Veglio, A.; Sereno, M.; Bussolino, F.; et al. Simplification of a complex signal transduction model using invariants and flow equivalent servers. *Theor. Comput. Sci.* **2011**, *412*, 6036–6057.
8. Pastor-Satorras, R.; Castellano, C.; Van Mieghem, P.; Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **2015**, *87*, 925–979.
9. Machens, A.; Gesualdo, F.; Rizzo, C.; Tozzi, A.E.; Barrat, A.; Cattuto, C. An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices. *BMC Infect. Dis.* **2013**, *13*, 185.
10. Sapiezynski, P.; Stopczynski, A.; Lassen, D.D.; Lehmann, S. Interaction data from the copenhagen networks study. *Sci. Data* **2019**, *6*, 1–10.
11. Huang, C.; Liu, X.; Sun, S.; Li, S.C.; Deng, M.; He, G.; Zhang, H.; Wang, C.; Zhou, Y.; Zhao, Y.; et al. Insights into the transmission of respiratory infectious diseases through empirical human contact networks. *Sci. Rep.* **2016**, *6*, 31484.
12. McCaw, J.M.; Forbes, K.; Nathan, P.M.; Pattison, P.E.; Robins, G.L.; Nolan, T.M. McVernon, J. Comparison of three methods for ascertainment of contact information relevant to respiratory pathogen transmission in encounter networks. *BMC Infect. Dis.* **2010**, *10*, 166.
13. Salathe, M.; Kazandjieva, M.; Lee, J.W.; Levis, P.; Feldman, M.W.; Jones, J.H. A high-resolution human contact network for infectious disease transmission. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 22020–22025.
14. Immuni. Available online: <https://www.immuni.italia.it> (accessed on 29 October 2020).
15. Großmann, G.; Backenköhler, B.; Wolf, V. Importance of Interaction Structure and Stochasticity for Epidemic Spreading: A COVID-19 Case Study. In *Quantitative Evaluation of Systems. QEST 2020. Lecture Notes in Computer Science, vol 12289*; Gribaudo M., Jansen D.N., Remke A., Eds.; Springer: Chambers, UK, 2020; pp. 211–219.
16. Chen, Y.C.; Lu, P.E.; Chang, C.-S.; Liu, T.-H. A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons. *arXiv* **2020**, arXiv:2003.00122.
17. Bobbio, A.; Cerotti, D.; Gribaudo, M.; Iacono, M.; Manini, D. Markovian Agent Models: A Dynamic Population of Interdependent Markovian Agents. In *Seminal Contributions to Modelling and Simulation. Simulation Foundations, Methods and Applications*; Al-Begain, K., Bargiela, A., Eds.; Springer: Chambers, UK, 2016; pp. 185–203.
18. Gribaudo, M.; Iacono, M.; Levis, A.H. An IoT based monitoring approach for cultural heritage sites: The Matera case. *Concurr. Comput. Pract. Exper.* **2017**, *29*, e4153.
19. Italian Government. Protezione Civile. 2020. Available online: www.protezionecivile.gov.it (accessed on 29 October 2020).
20. Corriere della Sera. Available online: https://www.corriere.it/cronache/20_agosto_04/coronavirus-veri-numeri-contagio-italia-15-milioni-casi-429ab770-d5a7-11ea-9379-bda036273e46.shtml (accessed on 20 August 2020).
21. Kurtz, T.G. Solutions of ordinary differential equations as limits of pure jump markov processes. *J. Appl. Probab.* **1970**, *7*, 49–58.
22. Bobbio, A.; Gribaudo, M.; Telek, M. Analysis of large scale interacting systems by mean field method. In Proceedings of the 5th International Conference on Quantitative Evaluation of Systems QEST, St Malo, France, 14–17 September 2008.