An Agent-Based Model of COVID-19 Diffusion to Plan and Evaluate Intervention Policies

Gianpiero Pescarmona, Pietro Terna, Alberto Acquadro, Paolo Pescarmona, Giuseppe Russo, Emilio Sulis, and Stefano Terna

Abstract A model of interacting agents, following plausible behavioral rules into a world where the Covid-19 epidemic is affecting the actions of everyone. The model works with (i) infected agents categorized as symptomatic or asymptomatic and (ii) the places of contagion specified in a detailed way. The infection transmission is related to three factors: the characteristics of both the infected person and the susceptible one, plus those of the space in which contact occurs. The model includes the structural data of Piedmont, an Italian region, but we can easily calibrate it for other areas. The micro-based structure of the model allows factual, counterfactual, and conditional simulations to investigate both the spontaneous or controlled development of the epidemic.

The model is generative of complex epidemic dynamics emerging from the consequences of agents' actions and interactions, with high variability in outcomes and stunning realistic reproduction of the successive contagion waves in the

G. Pescarmona · A. Acquadro · E. Sulis
University of Torino, Torino, Italy
e-mail: gianpiero.pescarmona@unito.it; alberto.acquadro@unito.it; emilio.sulis@unito.it

P. Terna (✉)
University of Torino, Torino, Italy

P. Terna (✉)
Fondazione Collegio Carlo Alberto, Turin, Italy
e-mail: pietro.terna@unito.it

P. Pescarmona
University of Groningen, Groningen, The Netherlands
e-mail: p.p.pescarmona@rug.nl

G. Russo
Centro Einaudi, Torino, Italy
e-mail: russo@centroeinaudi.it

S. Terna
TomorrowData, Turin, Italy
e-mail: stefano.terna@tomorrowdata.io
https://tomorrowdata.io

N. Bellomo, L. Gibelli (eds.), *Crowd Dynamics, Volume 3*, Modeling and Simulation in Science, Engineering and Technology, https://doi.org/10.1007/978-3-030-91646-6_9

reference region. There is also an inverse generative side of the model, coming from the idea of using genetic algorithms to construct a meta-agent to optimize the vaccine distribution. This agent takes into account groups' characteristics—by age, fragility, work conditions—to minimize the number of symptomatic people.

AQ1

OK, all the affiliations are fine

1 A Quick Introduction to Our Agent-Based Epidemic Model

The starting point is a compartmental model with Susceptible, Infected, and Recovered people (S.I.R.), but adding both a more detailed breakdown of the subjects involved in the contagion process [1] and a multi-scale framework to account for the interaction at different dimensional, and spatial levels [2]. From the virus micro-level, we move to individuals and up to the collective behavior of the population.

Following [3], we know that the analysis based on the assumption of heterogeneity strongly differs from S.I.R. compartmental structures modeled by differential equations. The authors of this work argue when it is best to use agent-based models and when it would be better to use differential equation models ponder when it is better to use agent-based models and when it would be better to use differential equation models. Differential equation models assume homogeneity and perfect mixing of characteristics within compartments, while agent-based models can capture heterogeneity in agent attributes and the structure of their interactions. We follow the second approach (about agent-based approach, see Sect. 1.1).

- Our model takes into consideration:

  (i) infected agents categorized as symptomatic or asymptomatic and
  (ii) the places of contagion specified in a detailed way, thanks to agent-based modeling capabilities.

- The infection transmission is related to three factors: the infected person's characteristics and those of the susceptible one, plus those of the space in which a contact occurs.

Finally, we subscribe the call of [4] to «cover the full behavioural and social complexity of societies under pandemic crisis» and we work arguing that «the study of collective behavior must rise to a "crisis discipline" just as medicine, conservation, and climate science have, with a focus on providing actionable insight to policymakers and regulators for the stewardship of social systems», as in [5].

A look at the structure of the whole presentation. In Sect. 1.1, we discuss models and specifically agent-based models; in Sect. 1.2, the molecular support to agents' intrinsic susceptibility construction; in Sect. 1.3, the structure of the model, with the daily sequence of the agents' actions. Section 2 introduces a detailed description of the internal model mechanisms, with: conditional actions in Sect. 2.1, parameters in Sect. 2.2 and agents' interaction in Sect. 2.3.

A technique for contagion representation is introduced in Sect. 3. Then we explore simulation cases in Sect. 4, building several batches of runs and comparing extreme situations in Sect. 4.1.

Section 4.2 reports the actual epidemic data in the reference region. With those data, we verify factual and counterfactual analyses in Sect. 5. Considering the possibility of calculating infection indicators without delays (Sect. 5.4), we experiment with the effect of adopting the control measure with 20 days of anticipation (Sect. 5.5). In Sect. 5.6 we verify another counterfactual policy, that of concentrating the efforts uniquely in defense of fragile persons. Section recap 5.7 summarizes these results.

The final application of the model is dedicated to a planning exercise on vaccination campaigns (Sect. 7). We introduce an analysis of the vaccine mechanism in the perspective of our model (Sect. 7.1), using both planned strategies (Sects. 7.4 and 7.5) and genetic algorithms (Sect. 7.6). The GAs goal is to optimize the behavior of a meta-agent, deciding the sequence of the vaccinations.

1.1 Why Models? Why Agents? Why Another Model?

Why another model, and most of all, why models? With [6]:

> The choice (...) is not whether to build models; it's whether to build explicit ones. In explicit models, assumptions are laid out in detail, so we can study exactly what they entail. On these assumptions, this sort of thing happens. When you alter the assumptions that is what happens. By writing explicit models, you let others replicate your results.

With even more strength:

> I am always amused when these same people challenge me with the question,"Can you validate your model?" The appropriate retort, of course, is,"Can you validate yours?" At least I can write mine down so that it can, in principle, be calibrated to data, if that is what you mean by "validate" a term I assiduously avoid (good Popperian that I am).

To reply to "why agents?", with [7] we define in short what an agent-based model is:

> An agent-based model consists of individual agents, commonly implemented in software as objects. Agent objects have states and rules of behavior. Running such a model simply amounts to instantiating an agent population, letting the agents interact, and monitoring what happens. That is, executing the model—spinning it forward in time—is all that is necessary in order to "solve" it.

More in detail:

> There are, ostensibly, several advantages of agent-based computational modeling over conventional mathematical theorizing. First, [...] it is easy to limit agent rationality in agent-based computational models. Second, even if one wishes to use completely rational agents, it is a trivial matter to make agents heterogeneous in agent-based models. One simply instantiates a population having some distribution of initial states, e.g., preferences. That is, there is no need to appeal to representative agents. [...] Finally, in most social processes either physical space or social networks matter. These are difficult to account for

> mathematically except in highly stylized ways. However, in agent-based models it is usually quite easy to have the agent interactions mediated by space or networks or both.

In [8] we have a relevant step ahead, considering *inverse generative social science*:

> The agent-based model (ABM) is the principal scientific instrument for understanding how individual behaviors and interactions, the micro-world, generates change and stasis in macroscopic social regularities. So far, agents have been iterated forward to generate such explananda as settlement patterns, scaling laws, epidemic dynamics, and many other phenomena [6]. But these are all examples of the forward problem: we design agents and grow the target phenomenon. The motto of generative social science is: "If you didn't grow it, you didn't explain it." [9] But there may be many ways to grow it! How do we find 'all' the non-trivial generators? This is inverse generative social science—agent architectures as model outputs not model inputs—and machine learning can enable it.

And now, "why another?" As a commitment to our creativity, using our knowledge to understand what is happening. Indeed, with arbitrariness: it is up to others and time to judge.

As any model, also this one is based on assumptions: time will tell whether these were reasonable hypotheses. Modeling the Covid-19 pandemic requires a scenario and the actors. As in a theater play, the author defines the roles of the actors and the environment. The characters are not real, they are prebuilt by the author, and they act according to their peculiar constraints. If the play is successful, it will run for a long time, even centuries. If not, we will rapidly forget it. Shakespeare's Hamlet is still playing after centuries, even if the characters and the plot are entirely imaginary. The same holds for our simulations: we are the authors, we arbitrarily define the characters, we force them to act again and again in different scenarios. However, in our model, the micro-micro assumptions are not arbitrary but based on scientific hypotheses at the molecular level, the micro agents' behaviors are modeled in an explicit and realistic way. In both plays and simulations, we compress the time: a whole life to two or three hours on the stage. In a few seconds, we run the Covid-19 pandemic spread in a given regional area.

1.2 The Molecular Basis of SARS-CoV-2 Infection

To fully understand what the word infection means, we have previously to define the scenario where life takes place.

We start with the properties of life on earth's surface [10]. Making a long story short, basically life is a dissipative process fueled by energy supplied by the sun. As the sun has been shining for billions of years the biological systems on earth expanded exponentially in a finite environment, and they became limited in their growth due to the shortage of available atoms/molecules ("nutrients"). The competition for the limiting nutrients in each local environment ("niche") will locally drive the selection and will explain the complexity of the interactions of the different organisms in any environment, from the microscopic to the social level.

The ground, the soil, and the ponds are overcrowded with bacteria, algae, molds, insects, and so on. They help to keep the ground healthy and ready for cultivation. We too are populated by microorganisms, the gut, the mucosae, the skin. But when we feel healthy we do not realize they are there; but sometimes we do not feel well, we are sick. We have a disease, we need a culprit: somebody different from us, a virus, a bacterium, a protozoan that infected us.

In most cases, the same agent is shared by people surrounding us, but most of them are healthy, few are sick. The coexistence/cooperation between organisms sharing the same "niche" is the rule after billions of years of evolution. The disease is the exception. The asymptomatic infection is the rule, the symptomatic infection is the disease. The simplest explanation for the rise of the symptoms is the competition of different organisms for a limiting "nutrient."

In the case of the Herpesvirus family and man, the limiting "nutrient" is Iron. Virus Ribonucleotide reductase is an enzyme with an affinity for iron higher than human cells. Infected cells survive quite well until iron availability covers the needs of both host and virus. In the case of iron shortage (evaluated as the level of serum ferritin) the infected cells are forced to reduce heme synthesis, necessary for the respiratory chain, and hence ATP synthesis. Less ATP, loss of many cellular functions, symptoms. In our experience, most of the people seropositive to HSV had no symptoms, provided they had serum ferritin levels $\geq$90 ug/dl. The lower the ferritin level, the higher the frequency of the symptoms. The level of ferritin depends on genetic, dietary, environmental factors, explaining the variability of the clinical manifestations [11].

In the case of HSV, the virus metabolism is well known and studied for tens of years. In the case of SARS-CoV-2 our experience is in the range of months and the identification of the limiting "nutrient" is only speculative.

On the basis of the data collected up to now, Cysteine could be the most relevant. One of the co-authors here, G. Pescarmona, with other contributors, has recently developed a software able to easily compare the amino acids (AA) percentage and some selected ratios between couples of them, using Uniprot proteins repository as data source [12]. Using this software, it has been possible to compare the AA percentage in different tissues [13] demonstrating the limiting role of AA local availability on the synthesis of specific proteins.

From the beginning of the pandemic, it has been clear that ACE2 was the preferred ligand for the Spyke protein and that cells expressing it were the perfect host for the fast synthesis of viral protein [14]. Our working hypothesis is that the best host cell is one producing a protein with a similar AA percentage. From Fig. 1 we can extract the following info: most AA percentages of the viral proteins are similar, with the exception of Cysteine (lower) and Methionine and Tryptophan (higher) to the ACE2. Higher methionine associates with faster protein synthesis, higher Tryptophan with higher nucleic acid synthesis. A perfect environment for replication of both RNA and virus proteins. Expression of viral proteins with high Cysteine decreases free cysteine and therefore Glutathione (GSH) synthesis, with impaired antioxidant defense and increased ROS activity. Increased ROS activity has been one of the first well-identified mechanisms of viral infection [15] and their

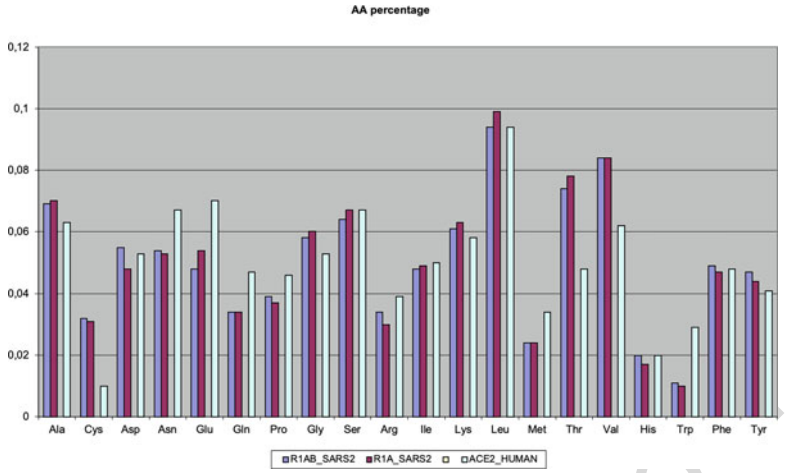


Fig. 1 Comparison of the Amino Acids percentage in the most representative SARS-Cov-2 proteins and the human ACE2, receptor on the surface of host cells

scavenging by GSH has been proposed as a preventive/therapeutic approach to the symptomatic disease [16].

Also, the ratio between AA couples, in Fig. 2, shows good similarity with the exception of the Spyke protein, as far as the ratios including glutamate are involved, but the almost perfect coincidence between the catalytic proteins of the virus and ACE2 explains the reproductive advantage of entering a cell expressing it. The interesting information that we get from this approach is that the cysteine deprivation of the naturally infected cells is shared also by cells induced to produce Spyke protein, independently by the vector used. Moreover, whilst the full virus enters the cells expressing on the outer surface ACE2, and we can identify them and try to imagine the long-term effects of infection, in the case of vaccines the synthetic vectors should allow the entry in any kind of cells.

In conclusion, we can expect oxidative damage (ROS increase and inflammation) in any kind of cell in our body. The extent of the inflammation will vary according to so many variables: age, diet, drugs, previous silent sites of inflammation, to make almost impossible the prediction of the localization and gravity of the side effects.

The Cytokine Storm

The cytokine storm is a synthetic definition of the set of reactions leading to the symptomatic COVID-19 and to death. The core process of the infection is the unbalance between ACE/ACE2. SARS-CoV-2 binds to ACE-2 and sequester it, causing an ACE prevalence and a sharp increase of ROS [17, 18]. All pre-existing processes leading to the prevalence of ACE are pro-inflammatory, those leading to a prevalence of ACE2, are anti-inflammatory. A low level of the active Vitamin D (1,25-dihydroxy-Vitamin D) leads to an increased expression of ACE. Cortisol has the same, but with an independent mechanism, effect on ACE expression and,

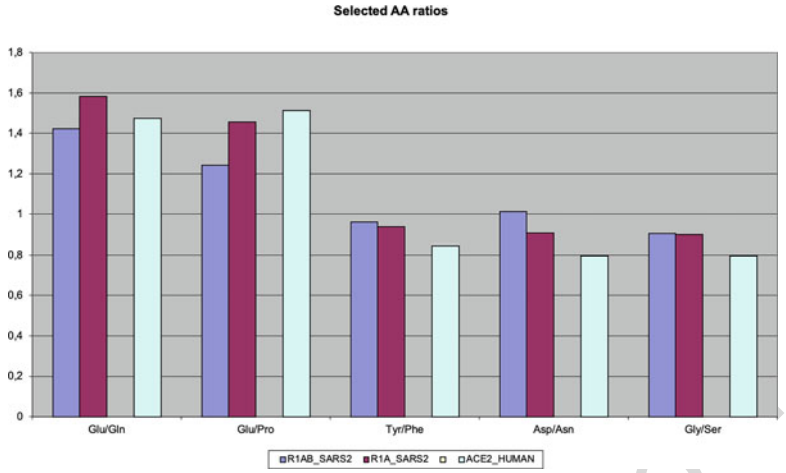


Fig. 2 Comparison of some ratios between selected Amino Acids in the most representative SARS-Cov-2 proteins and the human ACE2, receptor on the surface of host cells. These ratios supply specific information about the local metabolic condition inside the cell [13]

additionally, decreases the expression of ACE2. In all the cases the ROS released by ACE activity are scavenged by GSH and its ancillary enzymes. Downstream of ROS, the inflammatory pathway includes NF-kB, TNF-alpha, IL-6, PLA2, COX1, and COX2.

From the clinical point of view the COVID-19 pandemic is affecting differently the world population: in presence of conditions such as aging, diabetes, obesity, and hypertension the virus triggers a lethal cytokine storm and patients die from acute respiratory distress syndrome, whereas in many cases the disease has a mild or even asymptomatic progression [19]. The identification of the biochemical patterns underlying the severe disease may allow the identification of fragile people in need of more accurate protection.

Combining the biochemical determinants listed in Table 1 within the model described in Fig. 3 is possible to evaluate the risk for every individual, or class of similar individuals, of developing a severe form of the disease.

DHEA is an adrenal hormone, a precursor of testosterone and estrogens, that activates heme synthesis. Heme is required for plenty of reactions, including the respiratory chain (ATP synthesis) and Vitamin D activation. ATP is required for GSH synthesis, BMR reflects the activity of the respiratory chain and therefore depends again on heme. Heme synthesis requires iron, whose availability depends on diet, correct digestion, and absorption. Table 1 lists also some of the environmental factors that can interfere with the molecules involved in the COVID-19 dependent inflammatory response. Environmental pollution and drugs abuse in older people are among the factors that can explain the excess mortality in developed countries. The

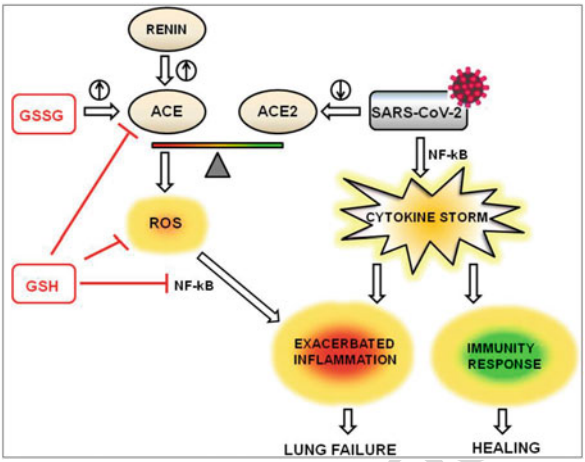


Fig. 3 All the main agents involved in the inflammatory response during COVID-19 are depicted here, with their relationships (reprinted with permission from [16])

Table 1 A synopsis of all the metabolic features associated with the clinical conditions favoring a severe COVID-19 development. DHEA: Dehydroepiandrosterone, GSH: Glutathione, Vit D: 25(OH)-Vitamin D, BMR: Basal Metabolic Rate

Risk factors	DHEA	Cortisol	GSH	Vit D	BMR	
Aging	Low	High	Low	Low	Low	t2.1 t2.2
Diabetes	Low	High	Low	Low	Low	t2.3
Hypertension	Low	High	Low	Low	?	t2.4
Obesity	Low	High	Low	Low	Low	t2.5
Diuretics	–	High	–	–	?	t2.6
Drugs	–	–	Low	Low	?	t2.7
Air pollution	–	–	Low	–	?	t2.8
Paracetamol	–	–	Low	–	?	t2.9
Chloroquine	–	–	Low	–	?	t2.10
Glucocorticoids	–	High	–	–	?	t2.11
Ibuprofen	–	–	–	–	?	t2.12
Aspirin						t2.13

therapeutic use of paracetamol, chloroquine, and glucocorticoids to prevent severe symptoms looks inappropriate on the basis of their action mechanism. 233 234

This set of considerations can be used to tentatively identify and protect fragile people but can be easily modified according to the epidemiological data. Unfortunately, up to now, the prevailing approach has been different, and not so much data about the characteristics of the patients severely ill have been published to allow validation of our criteria for fragility. 235 236 237 238 239

1.3 Our Model 240

With our model, we move from a macro compartmental vision to a meso and micro-analysis capability. Its main characteristics are: 241 242

- scalability: we take into account the interactions between virus and molecules inside the host, determining individual susceptibility; the interactions between individuals in more or less restricted contexts; the movement between different environments (home, school, workplace, open spaces, shops); the movements occur in different parts of the daily life, as in [20]; in detail, the scales are:
  - *micro*, with the internal biochemical mechanism involved in reacting to the virus, as in [16], from where we derive the critical importance assigned to an individual attribute of intrinsic susceptibility related to the age and previous morbidity episodes; the model indeed incorporates the medical insights and consistent perspectives of one of its co-authors, former full professor of clinical biochemistry, signing also the quoted article; a comment on Lancet [21] consistently signals the syndemic character of the current event: «Two categories of disease are interacting within specific populations—infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and an array of non-communicable diseases (NCDs)»;
  - *meso*, with the open and closed contexts where the agents behave, as reported above;
  - *macro*, with the emergent effects of the actions of the agents;
- granularity: at any level, the interactions are partially random and therefore the final results will always reflect the sum of the randomness at the different levels; changing the constraints at different levels and running multiple simulations should allow the identification of the most critical points, where to focus the intervention.

Summing up, S.I.s.a.R. (https://terna.to.it/simul/SIsaR.html) is an agent-based model designed to reproduce the diffusion of the COVID-19 using agent-based modeling in NetLogo [22]. We have Susceptible, Infected, symptomatic, asymptomatic, and Recovered people: hence the name S.I.s.a.R. The model works on the structural data of Piedmont, an Italian region, but we can quite easily calibrate it for other areas. It reproduces the events following a realistic calendar (national or local government decisions, as in Sect. 2.2), via its script interpreter. At the above address, it is also possible to run the code online without installation. Into the *Info* sheet of the model, we have more than 20 pages of Supporting Information about both the structure and the calibration of the model.

The micro-based structure of the model allows factual, counterfactual, and conditional simulations. Examples of counterfactual situations are those considering:

(i) different timing in the adoption of the non-pharmaceutical containment measures;
(ii) an alternative strategy, focusing exclusively on the defense of fragile people.

The model generates complex epidemic dynamics, emerging from the consequences of agents' actions and interactions, with high variability in outcomes, and with a stunning realistic reproduction of the contagion waves that occurred in the reference region.

We take charge of the variability of the epidemic paths within the simulation, running batches of executions with 10,000 occurrences for each experiment.

Following [8], the AI and inverse generative side of the model comes from constructing a meta-agent optimizing the vaccine distribution among people groups—characterized by age, fragility, work conditions—to minimize the number of symptomatic people (as deceased persons come from there).

We can characterize the action of the planner both:

(i) introducing ex-ante rules following "plain" or "wise" strategies that we imagine as observers or
(ii) evolving those strategies via the application of a genetic algorithm, where the genome is a matrix of vaccination quotas by people groups, with their time range of adoption.

2 How S.I.s.a.R. Works

We have two initial infected individuals in a population of 4350 individuals, on a scale of 1:1000 with Piedmont. The size of the initial infected group is out of scale: it is the smallest number ensuring the epidemic's activation in a substantial number of cases. Initial infected people bypass the incubation period. For plausibility reasons, we never choose initial infected people among persons in nursing homes or hospitals. The presence of agents in close spaces—such as classrooms, factories, homes, hospitals, nursing homes—is set with realistic numbers, out of scale: e.g., a classroom contains 25 students, a home two persons, large factories up to 150 employees, small ones up to 15, etc.; the movements occur in different parts of the daily life, as in [20].

In Fig. 4 we have a 3D representation of the model world, with one of the possible random maps that the simulation generates. Persons are in gray, houses in cyan, nursing homes in orange, hospitals in pink, schools in yellow, factories (with shops and offices) in brown. Persons have a cylinder as shape, if regular or robust (young); a capital X if fragile; temporary their colors can be: red, if symptomatic; violet, if asymptomatic; turquoise, if symptomatic recovered; green, if asymptomatic recovered.

Doing the batches of repetitions of the simulation, we use random maps to have a neutral effect of the structure of the space.

We can set:

- min and max duration of the individual infection;
- the length of the incubation interval;
- the critical distance, i.e., the radius of the possibility of infection in open air, with a given probability;
- the corrections of that probability, due to the personal characteristics of both active and the passive agents;

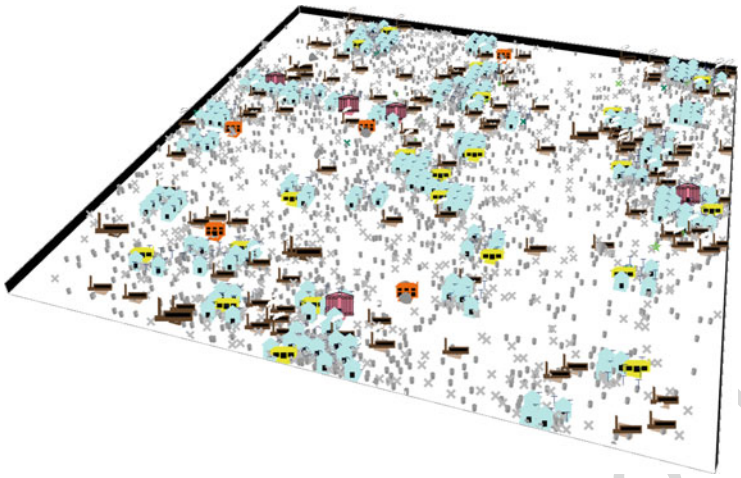


Fig. 4 A live 3D picture of the model world

- active agents can be symptomatic or asymptomatic, with different spreading characteristics (see (ii) in Sect. 2.2);
- passive agents, as receivers, can be robust (young), regular, fragile, and extra fragile.

We have two main types of contagion: (a) within a radius, for people moving around, temporary in a house/factory/nursing home/hospital; (b) in a given space (room or apartment) for people resident in their home or in a hospital or in a nursing home or being in school or in a working environment.

People in hospitals and nursing homes can be infected in ways (a) and (b). Instead, while people are at school, they can only receive the disease from people in the same classroom, where only teachers and students are present, so this is a third infection mechanism (c). In all cases, the personal characteristics of the recipients are decisive.

We remark that workplaces are open to all persons, as clients, vendors, suppliers, external workers can go there. In contrast, schools are reserved for students and school operators.

All agents have their home, inside a city, or a town. The agents also have usual places (UPs) where they act and interact, moving around. These positions can be interpreted as free time elective places. When we activate the schools, students and teachers have both UPs and schools; healthcare operators have both UPs and hospitals or nursing homes; finally, workers have both UPs and working places. In each day (or tick of the model), we simulated full sequences of actions.

Figure 5 describes what happens during every *day* in our simulated world, with the daily sequences of actions.

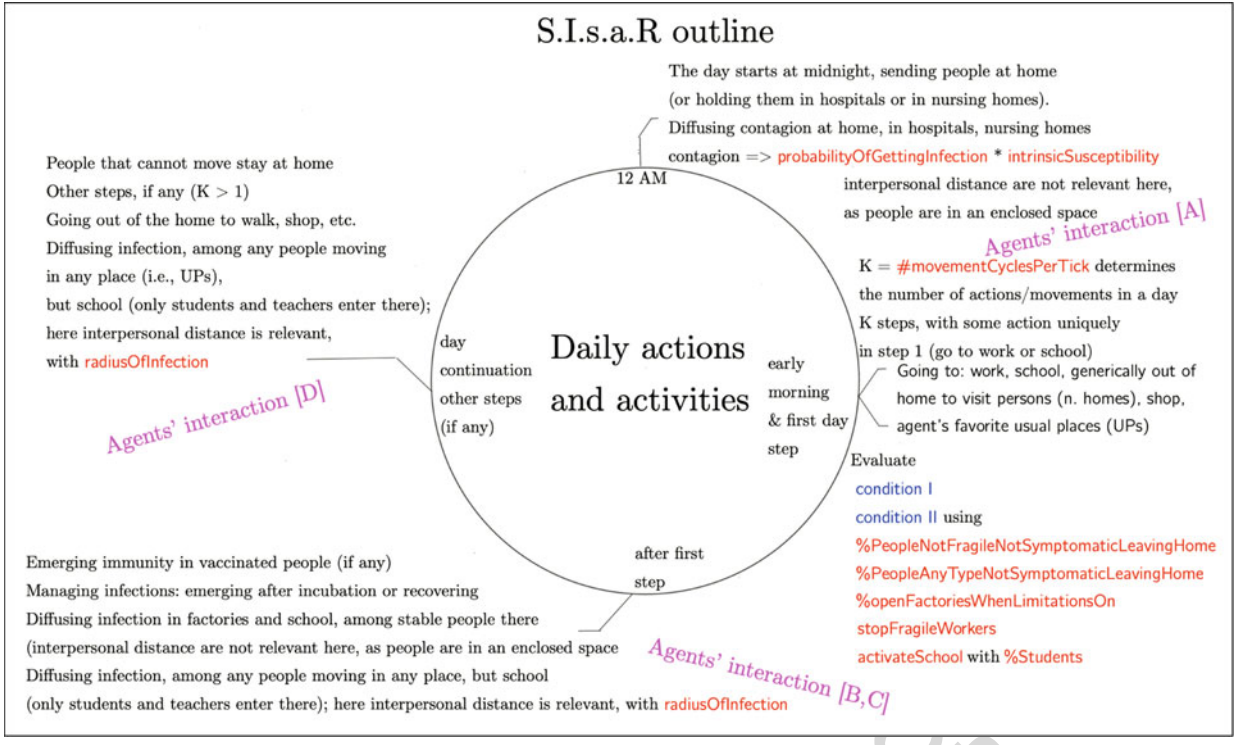


Fig. 5 A day in the simulation, with N repetitions where N is the duration of a given outbreak; look at: Sect. 2.1 for the rules of the conditional actions; Sect. 2.2 for the parameter definitions; and Sect. 2.3 for details on the agent interactions

2.1 Conditional Actions

Agents' movements in space, to go to work, school, and other UPs are subject to two interrelated general conditions.

I Symptomatic persons are at home or in a hospital or a nursing home and do not move.

II People not constrained by *condition I* can move if (primary rule) there are no general limitations (e.g., lockdown) *OR* if one of the following sub-conditions applies:

(a) agents who are hospital healthcare operators or nursing home healthcare operators;
(b) all people, according to the probability of moving of the whole non-symptomatic agents (Sect. 2.2, (iv));
(c) regular people, according to the probability of moving of the regular non-symptomatic agents (Sect. 2.2, (v));
(d) workers, if all the factories are open or it is open their own workplace (Sect. 2.2, (vi));
(e) teachers, if the schools are open (Sect. 2.2, (vii));
(f) students, if the schools are open, but with a possible quota limitation (Sect. 2.2, (viii)).

2.2 Parameter Definition

We define the parameters of Fig. 5, also with their short names used in program scripts, in round brackets. The values of the parameters are reported in detail in Appendix 1—Parameter values (Sect. 9).

(i) *probabilityOfGettingInfection* (`prob`) is the base probability of getting infected, to be multiplied by the *intrinsicSusceptibility* factor (iii); it is activated if the subject is within a circle of radius (ix) with an infected person; values at (Sect. 9, (i));
(ii) *D%*, without the short name, is the percent increasing or decreasing factor of the contagion spread of an asymptomatic subject, compared to that of a symptomatic one, value at (Sect. 9, (ii));
(iii) the *intrinsicSusceptibility* in defined in Eq. (1)

$$intrinsicSusceptibility = intrinsicSusceptibilityFactor^{groupFragility} \tag{1}$$

with *intrinsicSusceptibilityFactor* set to 5, and $groupFragility$ exponent set to:

1 for extra-fragile persons,
0 for fragile persons,
−1 for regular persons,
−2 young people from 0 to 24 years old;

(iv) *%PeopleAnyTypeNotSymptomaticLeavingHome* (`%PeopleAny`) determines, in a probabilistic way, the number of people of any kind going around in case of limitations/lockdown; the limitations operate only if the lockdown is on (into our simulated world, from day 20); values at (Sect. 9, (iv));

(v) *%PeopleNotFragileNotSymptomaticLeavingHome* (`%PeopleNot`) determines, in a probabilistic way, the number of regular people going around in case of limitations/lockdown; as above, the limitations operate only if the lockdown is on (into our simulated world, from day 20); values at (Sect. 9, (v)); we try to reproduce the uncertainty of the decisions in the real world into the model via frequent changes of the parameters (iv) and (v);

NB, the parameters (iv) and (v) produce independent effects, as in the following examples: (a) the activation of *%PeopleAny at 31, 0* and, simultaneously, of *%PeopleNot at 31, 80*, means that people had to stay home on that day, but people specifically not fragile could go out in 80% of the cases; (b) *%PeopleAny at 339, 80* and, simultaneously, *%PeopleNot at 339, 100* means that fragile and not fragile persons cannot always go around, but only in the 80% of the cases; instead, considering uniquely non-fragile persons they are free to go out; the construction is an attempt to reproduce a fuzzy situation; in future versions of the model, we will define the quotas straightforwardly:

- `%FragilePeopleNotSymptomaticLeavingHome`;
- `%NotFragilePeopleNotSymptomaticLeavingHome`;

(vi) *%openFactoriesWhenLimitationsOn* (`%Fac`) determines, in a probabilistic way, the factories (small and large industries, commercial surfaces, private and government offices) that are open when limitations are on; if the factory of a worker is open, the subject can go to work, not considering the restrictions (but uniquely in the first step of activity of each day); values at (Sect. 9, (vi));

(vii) *stopFragileWorkers* (`sFW`) is *off* (set to 0) by default; if *on* (set to 1), fragile workers (i.e., people fragile due to prior illnesses) can move out of their homes following the (iv) and (v) parameters, but cannot go to work; in the *off* case, workers (fragile or regular) can go to their factory (if open) also when limitations are on; values at (Sect. 9, (vii)); alternatively, we also have the `fragileWorkersAtHome` parameter; if *on* (set to 1) the total of the workers is unchanged, but the workers are all regular; we can activate this counterfactual operation uniquely at the beginning of the simulation;

(viii) when *activateSchools* (`aSch`) is *on* (set to 1), teachers and students go to school avoiding restrictions (but uniquely in the first step of activity of each day); *%Students* (`%St`) sets the quota of the students moving to school; the residual part is following the lessons from home; values at (Sect. 9, (viii));

(ix) following *radiusOfInfection* (`radius`), the effect of the contagion—outside enclosed spaces, or there, but for temporary presences—is possible within that distance; values at (Sect. 9, (ix));

(x) *asymptomaticRegularInfected%* and *asymptomaticFragileInfected%* are the parameters determining the percentage of asymptomatic persons after a contagion for non-fragile (all cases) or fragile people; they are without short names, as they come directly from the model interface; we can see the interface online, activating the model at https://terna.to.it/simul/SIsaR.html; values at (Sect. 9, (x)).

2.3 *Agents' Interaction*

We underline that our simulation tool is not based on micro-simulation sequences, calculating the contagion agent by agent, on the base of their characteristics and ex-ante probabilities. It implements a true agent-based simulation, with the agents acting and, most of all, interacting. The effect is that of generating continuously contagion situations.

Each run creates a population with expected characteristics, but also with random specifications, to assure the heterogeneity in agents. The daily choices of the agents are partially randomized, to reproduce real-life variability.

Contagions arise from agents' interactions, in four time phases, as specified in Fig. 5:

A in houses (at night), hospitals, nursing homes;
B in schools and workplaces in general, among people stable there;
C in the places above (excluding schools) by people temporary there and in open spaces (UPs above);
D interactions mainly in open spaces (UPs above).

3 Contagion Representation

We introduce a tool analyzing the contagions' sequences in simulated epidemics and identifying the places where they occur.

- We represent each infected agent as a horizontal segment (from the starting date to the final date of the infection) with vertical connections to other agents receiving the disease from it.
- We represent the new infected agents via further segments at an upper level.
- We display multiple information using three elements.

  - Colors in horizontal segments (areas of the infections): black for unknown places, gray for open spaces, cyan for houses, orange for nursing homes, pink for hospitals, yellow for schools, brown for factories, with shops and offices.
  - Vertical connecting segments keep the same color of the horizontal generating one.
  - Line thickness; proportional to fragility.
  - Styles: dotted lines for incubation, dashed lines for asymptomatic subjects, solid lines of symptomatic ones.

- This graphical presentation enables understanding at a glance how an epidemic episode is developing. In this way, it is easier to reason about countermeasures and, thus, to develop intervention policies.

At https://github.com/terna/contagionSequence we have the program *sequentialRecords.ipynb*, generating these sequences.

Figure 6 is useful as an example. We start with two agents from the outside, with black as the color code (unknown place). The first one is young—as reported by the thickness of the segment, with the infection starting at day 0 and finishing at day 22—and asymptomatic (dashed line); it infects no one. The second one—regular, as reported by the thickness of the segment, with the infection starting at day 0 and finishing at day 15—is asymptomatic (dashed line) and infects four agents on day 2. All the four infected agents receive the infection at work (brown color) and turn to be asymptomatic after the days of incubation (dotted line); the first and the fourth are regular agents; the second and the third are fragile ones.

Continuing the analysis: on day 3, the second agent infects three other agents (at home, at work, at work) [...]; on day 13, agent number five infects seven regular agents at work and an extra-fragile one in a nursing home (orange color), etc.

If a vertical segment changes its color, we have an agent in an upper layer infecting someone on the same day of the infection transmitted by an agent in a lower row, so we lose some graphical information.

In Fig. 7 we see the example of an epidemic with non-pharmaceutical containment measures in adoption: a first wave shows an interlaced effect of contagions at home, in nursing homes, and at work. After a phase in which contagions develop mainly at home, a skinny bridge connects the first wave to a second one, which restarts from workplaces. The thickness of the *snake* of the contagions measures the stock on infects agents on a given date; the slope reports the speediness of the epidemic development; the upper vertical coordinate reports the cumulative number of infected people.

In Appendix 2—A gallery of contagion sequences (Sect. 10), we have several examples of contagion sequences.

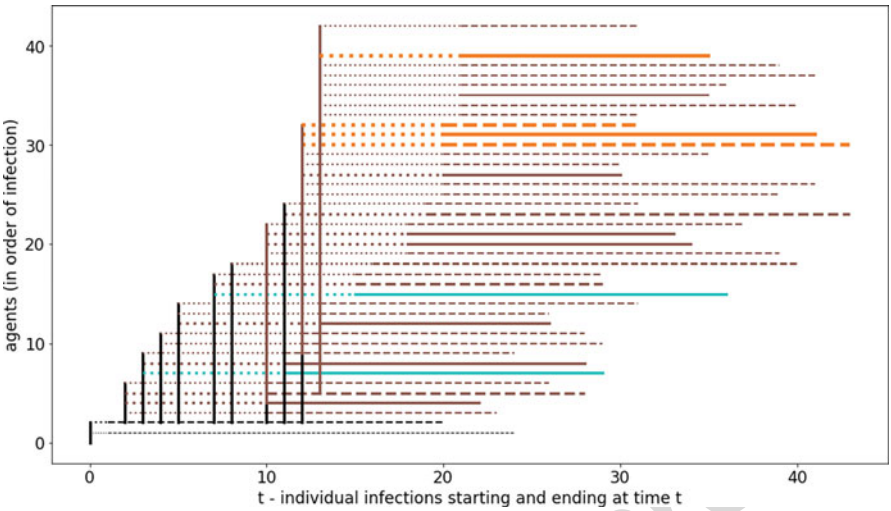


Fig. 6 A case with containment measures, first 40 infections: workplaces (brown) and nursing homes (orange) interweaving

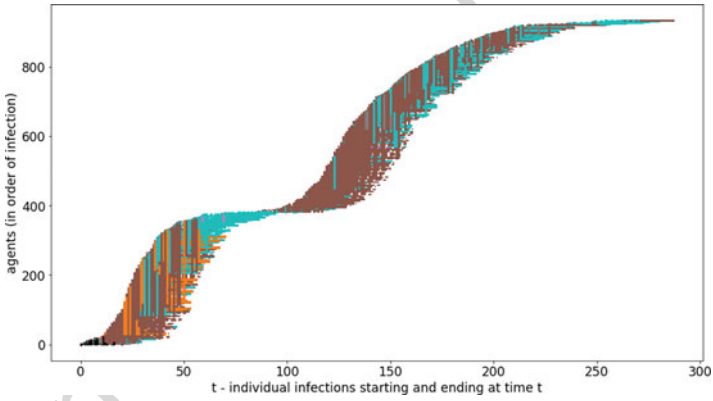
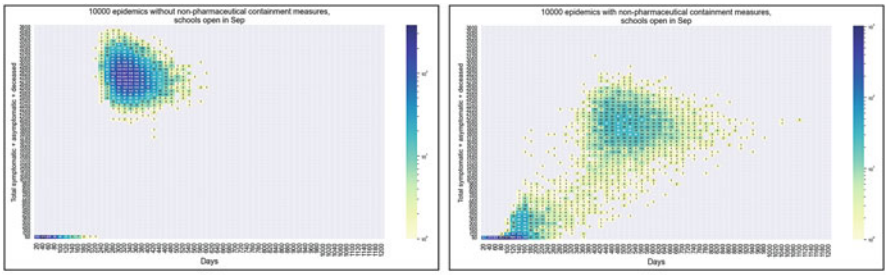


Fig. 7 A case with containment measures, the whole epidemics: workplaces (brown) and nursing homes (orange) and then houses (cyan), with a bridge connecting two waves

4 Exploring Scenarios with Simulation Batches

The sequences described in Sect. 3 ~~offer~~ suggest possible interventions, but are single cases. To explore systematically the introduction of factual, counterfactual, and prospective actions, we need to analyze batches of simulations. In this perspective, each simulation run—whose length coincides with the disappearance of symptomatic or asymptomatic contagion cases—is a datum in a set of different duration and contagion outcomes. To compare the consequences of each batch's



(a) (b)

Fig. 8 Starting our analyses: 10,000 epidemics in Piedmont. (**a**) Outbreaks without non-pharmaceutical containment measures. (**b**) Outbreaks with non-pharmaceutical containment measures

basic assumptions, we need to represent compactly the results emerging from simulation repetitions.

We use blocks of ten thousand repetitions. Besides summarizing the results with the usual statistical indicators, we adopt the technique of the heat-maps. With [23], our goal is that of making comparative analyzes, not forecasts. This consideration is consistent with the enormous standard deviation values that are intrinsic to the specific reality.

At https://github.com/terna/readSIsaR_BatchResults we have the codes producing the maps of the batches. A heat-map is a double histogram: in our application, it displays each simulated epidemic's duration in the x axis and the total number of the symptomatic, asymptomatic, and deceased agents in the y axis (on a scale of 1:1000). Each cell contains the number of epidemics with x duration and y outcome. Besides the number, a logarithmic color scale improves the readability of the maps.

4.1 Epidemics Without and With Control Measures

As a starting point, we compare the situations represented in Fig. 8a, b. In Fig. 8a, the heat-map reports the distribution in duration and infection causation of 10,000 simulated outbreaks left to spread without any control; coherently, with the school always open. The results in Table 2 are scary. The concentration of the cases in the heat-map shows that, except a few instances spontaneously concluding in a short period (left bottom corner), produces a heavy *cloud* of cases lasting one year or one year and a half, hitting (as symptomatic, asymptomatic, and deceased) from 2000 to 3500 persons on a total of 4350 in the region (scale of 1:1000).

In Fig. 8b and the related Table 3, we report a similar simulation batch of 10,000 runs of the model, but with the adoption of the basic non-pharmaceutical containment measures, registered in the values of the parameters in Appendix 1— Parameter values (Sect. 9). A calendar is at https://terna.to.it/simul/calendario092.pdf,

Table 2 Mean values and standard deviations in Fig. 8a cases

(000)	Symptomatic	Totalinfected&Deceased	Duration	t4.1
Mean	969.46	2500.45	303.10	t4.2
Std	308.80	802.88	93.50	t4.3

Table 3 Mean values and standard deviations in Fig. 8b cases

(000)	Symptomatic	Totalinfected&Deceased	Duration	t6.1
Mean	344.22	851.64	277.93	t6.2
Std	368.49	916.41	213.48	t6.3

and the model—version 0.9.6—is updated until April 2021. The results are dramatically different, showing the efficacy of the containment measures.

4.2 Actual Data

The critical points for our simulation experiments in Piedmont are Summer and Fall 2020 in Fig. 9a, where we have the time series of the first part of Piedmont's actual epidemic. The blue line represents the cumulative number of infected persons. Initially, only symptomatic cases were accounted for, but after the 2020 Summer, with more generalized tests, also asymptomatic patients are included:

- from http://www.protezionecivile.it/web/guest/department, the Italian Civil Protection Department web site, we find at https://github.com/pcm-dpc/COVID-1, i.e., the repository of regional data;
- we observe data about symptomatic infected people in the first wave, but from October 2020, data are mixed: in the above *git* repository, in October and November, we had "Positive cases emerged from clinical activity," unfortunately then reported as "No longer populated" (from the end of November, our observation) and "Positive cases emerging from surveys and tests, planned at national or regional level," again "No longer populated" (from the end of November, our observation);
- as a consequence, the subdivision between symptomatic and asymptomatic cases is impossible after that date.

Considering the dynamic of the data in Fig. 9a, we search within the simulation batch for cases with both:

(i) numbers of infected persons quite similar at `cp2` and at `cp3`; besides, numbers not too different from those of the figure; (with *cp*, we indicate the internal check points of the simulation program; in Fig. 9a we also report the number of days from the beginning of the epidemic for each check point);
(ii) the number of infected persons at `cp4` has to be significantly greater than those at the previous check point.

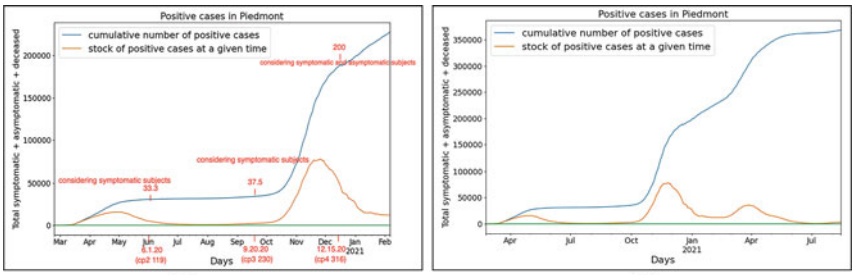


Fig. 9 Actual data. (**a**) Critical points in epidemic dynamic in Summer and Fall 2020 in Piedmont. (**b**) Data in Piedmont until July 2021, showing three waves

In a lot of cases, epidemics satisfying condition (i) fail to match condition (ii); both the situations happen only in less than the 1.5% of the instances in a batch of ten thousand epidemic. We can guess that the second wave registered in Piedmont after the Summer "pause" is due to new infected agents coming from outside and restarting the contagion process.

Other critical points in our analysis are the day on which the vaccination campaign starts, 373 of the simulation (Feb. 12th, 2021), and the day of the effectiveness of the initial vaccinations, 40 days later, day 413 (Mar. 22nd, 2021). At those dates, within the simulations, we can find either the presence of many infected agents or of few ones, as effectively was the situation in Piedmont.

NB, we concluded model calculations in April 2021. In Fig. 9b, the time series covering the whole period.

5 Factual and Counterfactual Analyses

In Fig. 10 we collect the heat-maps of the experiments:

- observing the emergence of spontaneous second waves, in the absence of specific control measures (Sect. 5.1);
- causing the emergence of the second wave through infections from outside, again in the absence of specific control measures (Sect. 5.2);
- causing the emergence of the second wave through infections from outside, in the presence of specific control measures (Sect. 5.3);
- reproducing the case of Sect. 5.3, anticipating by twenty days the start and end of all control measures (Sect. 5.5);
- reproducing the case of Sect. 5.3, limiting the control measures to fragile workers and other fragile people (Sect. 5.6).

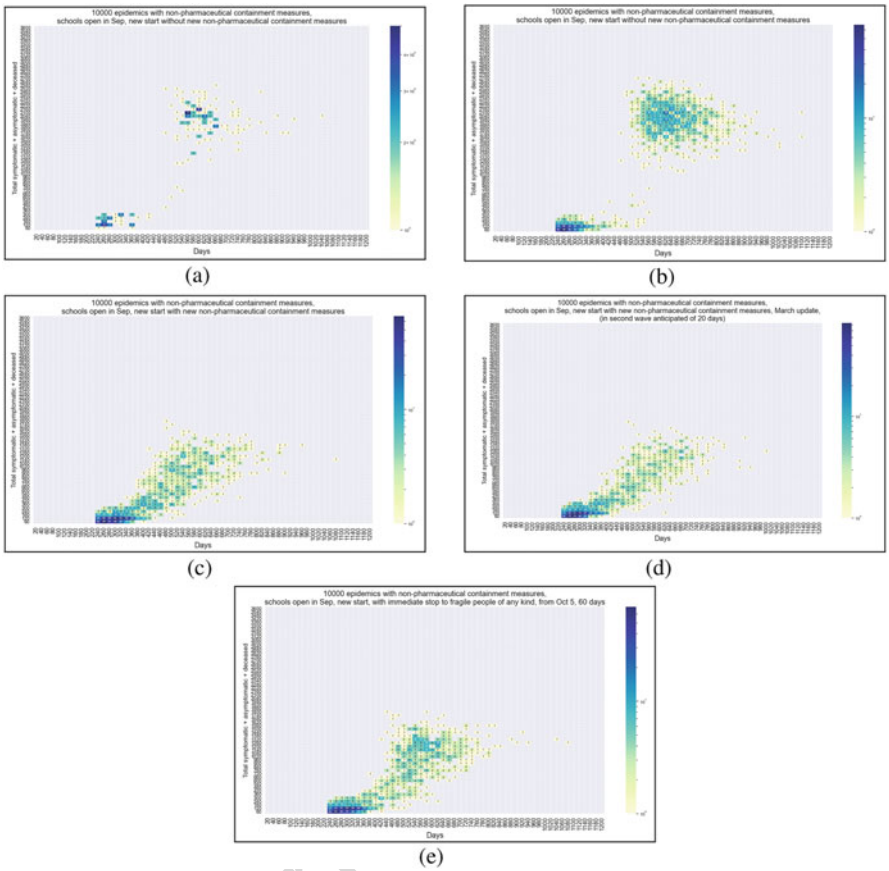


Fig. 10 Heat-maps of the factual and counterfactual analyses. (**a**) First wave with non-pharmaceutical containment measures, spontaneous second wave, without specific measures. (**b**) First wave with non-pharmaceutical containment measures, forcing the second wave, without specific measures. (**c**) First wave with non-pharmaceutical containment measures, forcing the second wave, *with new specific non-ph. containment measures*. (**d**) First wave with non-pharmaceutical cont. meas,, forcing the second w., *with new specific non-ph. cont. meas., acting 20 days in advance*. (**e**) First wave with non-ph. cont. meas., forcing the sec. wave; *in sec. wave, uniquely stopping fragile people, including fragile workers*

5.1 Spontaneous Second Wave, Without Specific Containment Measures

In an initial plain batch of runs of the Piedmont model, we count only 140 cases of epidemics with both the absence of new contagions in Summer 2020 and their explosion in Fall, as in Fig. 9a.

The steps are:

- we select, first of all, the 170 cases of epidemics that have, on June 1st, a number of symptomatic agents in the (10, 70] interval (with mean: 37.9) and, on September 20th, a number of symptomatic agents in the (20, 90] interval (with mean: 60.4);
- due to the lack of data described in Sect. 4.2, to compare December 15th and September 20th situations, we use symptomatic plus asymptomatic agents' count;
- we observe the existence of 140 outbreaks with the required characteristics; the December mean of the infected agents is 648.7, sensibly larger than the actual value: ≈200.0.

OK, thanks for the correction

AQ3

We overestimate the reality being the long-lasting simulated outbreaks, the larger ones, and, most of all, having no containment measures operating in the simulations.

Figure 10a and Table 4 show the outbreaks with similar cumulative numbers before and after the Summer 2020 "pause" (170 cases), with the second wave (140 cases) in the absence of containment measures.

140 out of 10,000, i.e., 1.4%, is a very light spontaneous ratio for the second wave occurred in the Fall. The transition to the third wave, that we see in Fig. 9b, is easy to explain, as the second wave never completely ended.

5.2 Second Wave, New Infections from Outside, Without Specific Containment Measures

To generate a framework consistent with the presence of a second wave after a period of substantial inactivity of the epidemic, we introduced two cases of infected persons coming back from outside after Summer vacancies, conventionally on September 1st, 2020.

As above, the steps are:

- we select, first of all, the 1407 cases of epidemics that. on June 1st have, a number of symptomatic agents in the (10, 70] interval (with mean: 35.6) and, on September 20th, a number of symptomatic agents in the (20, 90] interval (with mean: 40.0);
- due to the lack of data described in Sect. 4.2, to compare December 15th and September 20th situations, we use symptomatic plus asymptomatic agents' count;
- we observe the existence of 1044 outbreaks with the required characteristics; the December mean of the infected agents is 462.1, again sensibly larger than the actual value: ≈200.0.

We overestimate the reality being the simulations run without the adoption of containment measures.

Both Fig. 10b and Table 5 show the outbreaks with similar cumulative numbers before and after the Summer 2020 "pause" (1407 cases), with the second wave of

Table 4 Spontaneous second wave, without specific measures

t9.1	(1000)	Jun 1, 20		Sep 20, 20		Dec 15, 20		Feb 1, 21		May 1, 21		Dec 15, 20 to end		
t9.2	Cum. v.	Sym.	All	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	Totalinf.	days
t9.3	Count	170.0	170.0	170.0	170.0	140.0	140.0	131.0	131.0	128.0	128.0	140.0	140.0	140.0
t9.4	Mean	37.9	100.2	60.4	159.3	248.4	648.7	432.2	1109.5	656.3	1655.5	701.1	1757.9	594.2
t9.5	Std	16.4	61.0	19.6	71.7	167.4	424.3	220.4	538.4	215.4	513.3	246.4	599.7	118.9

Table 5 Second wave, new infections from outside, without specific measures

t12.1	(1000)	Jun 1, 20		Sep 20, 20		Dec 15, 20		Feb 1, 21		May 1, 21		Dec 15, 20 to end		
t12.2	Cum. v.	Sym.	All	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	Totalinf.	Days
t12.3	Count	1407.0	1407.0	1407.0	1407.0	1044.0	1044.0	1005.0	1005.0	980.0	980.0	1044.0	1044.0	1044.0
t12.4	Mean	35.6	72.7	40.0	84.1	180.4	462.1	354.1	900.4	623.8	1563.3	726.6	1810.9	620.9
t12.5	Std	14.1	42.6	16.7	52.8	134.6	354.6	213.8	535.4	217.9	527.0	221.9	544.0	110.8

1044 cases. In the absence of containment measures, we have a heavy cloud as that of Fig. 8a, with infected people of any kind in a range approximately of 1500 to 2800 realizations, with an equivalence, to the Piedmont scale, to 1.5–2.8 millions of subjects.

The number of cases is now sufficient to evaluate the effects of factual (Sect. 5.3 and counterfactual (Sects. 5.5 and 5.6) simulation experiments.

5.3 *Second Wave, New Infections from Outside, with New Specific Containment Measures*

Repeating the third step above:

- we observe the existence of 874 outbreaks with the required characteristics; the December mean of the infected agents is 340.6, closer to the actual value (≈200.0) due to the introduction into the simulation of specific control measures for the second wave.

We always overestimate the reality because the surviving epidemics are the larger ones.

In Fig. 10c we see that the heavy cloud of the previous figure dissolved, and in Table 6 the numbers in italic emphasize the positive effects of the containment interventions on the cases of epidemic continuation (which have also dropped in quantity).

5.4 *Calculating the Reproduction Number Without Delays*

The reproduction number R_t [24, 25]

> is the average number of secondary cases of disease caused by a single infected individual over his or her infectious period

and is defined as follows:

$$R_t = \frac{I_t}{\sum_{s=1}^{t} w_s I_{t-s}}, \tag{2}$$

where:

- I_t is the number of new infected individuals at time t
- $w_s = \Gamma(s; \alpha, \beta)$ is the infectivity profile, usually approximated with the serial interval distribution [25]; it shapes the infectious period of each individual by weighting the infected individuals so that when their period is over, they do not

Table 6 Second wave, new infections from outside, with new specific measures

t15.1	(1000)	Jun 1, 20		Sep 20, 20		Dec 15, 20		Feb 1, 21		May 1, 21		Dec 15, 20 to end		
t15.2	Cum. v.	Sym.	All	Sympt.	TotalInf.	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	TotalInf.	Sympt.	Totalinf.	Days
t15.3	Count	1407.0	1407.0	1407.0	1407.0	874.0	874.0	719.0	719.0	523.0	523.0	874.0	874.0	874.0
t15.4	Mean	35.6	72.7	40.0	84.1	*130.0*	*340.6*	*194.4*	*512.8*	*295.7*	*791.2*	252.7	666.4	494.1
t15.5	Std	14.1	42.6	16.7	52.8	83.9	232.6	104.1	276.9	119.1	300.6	156.8	416.4	122.7

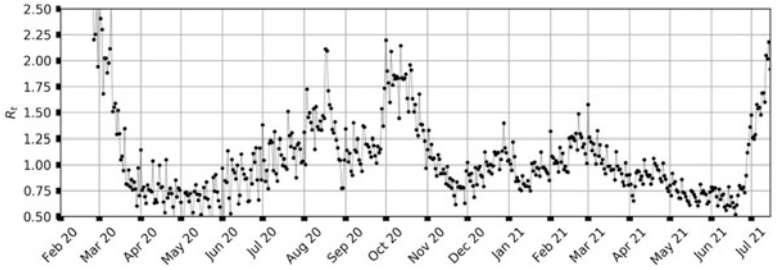


Fig. 11 Naive R_t calculated on raw infected cases by symptoms onset date, data-set by ISS

count any more in the sum; it is usually assumed to be the Gamma distribution [25]

- there is great uncertainty on the parameters of the Gamma distribution, which have been fitted to different values on different national data-sets ([26], table 1 page 25)
- following the Istituto Superiore di Sanità (ISS), italian R_t estimates are based on the parameters fitted in [27], namely $\alpha = 1.87$ and $\beta = 0.28$

While Eq. (2) could in principle be applied naively to any time series of the new infected cases, it usually leads to noisy results caused by the noise contained in the original raw series, as can be seen in Fig. 11. Moreover, despite the noisy content of the original signal, the naive approach does not give any clue on the confidence interval of the result, which is fundamental if the reproduction number has to be used to take decisions about the restrictions.

The most widely adopted approach to extract statistics about the R_t estimate, and hence its confidence interval, is to apply Bayesian statistical inference, assuming a prior distribution for the serial interval and a posterior for the reproduction number [25].

While Bayesian inference allows us to compute any kind of statistics on the estimate, it still fails dealing with the noise in the original signal, leading again to spiky estimates of R_t.

The standard solution to smooth out the noise is to assume that the transmissibility is constant over a time window (e.g., a week): we can then estimate the average R_t over the time window [25], by computing the total number of new infected cases over a window τ instead of those of each single day: $\hat{I}_{t,\tau} = \sum_{s=t-\tau}^{t} I_s$ and replacing it to I_t in Eq. (2); note that this is equivalent to compute R_t on the average of I_t over the window, as Eq. (2) is invariant under constant scaling of I_t.

The result is smoothed, but it turns out that it is delayed by the size of the windows. Figure 12 shows R_t calculated over a 14-day rolling window (14 days is the window size officially adopted in Italy); it is clearly visible that the average R_t is systematically delayed: maximizing the cross correlation of the signals confirms a measure of the delay of 14 days.

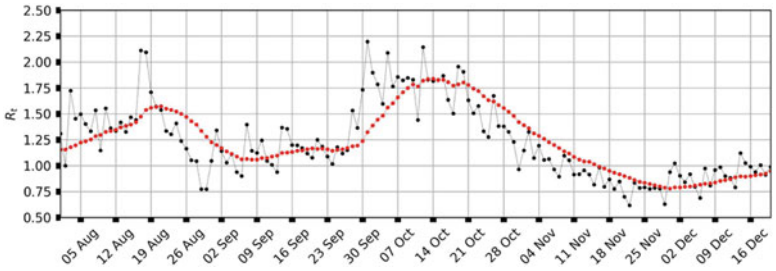


Fig. 12 In black R_t calculated on raw infected cases, in red the average R_t calculated over a window of 14 days; both series are by symptoms onset date, data-set by ISS

Official Data-Sets

Data used in all the computations refers to the following sources:

- data-set by ISS: count of new infected individuals by symptoms onset date, at https://github.com/tomorrowdata/COVID-19/blob/main/data/sources/ISS/covid_19-iss_2021-07-30T22:34:44%2B00:00.inizio_sintomi.csv downloaded on Jul 30
- data-set by Protezione Civile: count of new infected individuals by notification date, at https://github.com/pcm-dpc/COVID-19 downloaded on Jul 31

5.4.1 Tikhonov Regularization to Smooth the Original Signal

As an alternative solution to averages, we adopt Tikhonov regularization to the original signal, which does not introduce delays.

I_t is smoothed by fitting a series to represent the derivative of I_t and then integrating it back to the original signal, which then results in a smoothed one.

We search for the differential signal $\boldsymbol{\omega}$ such that:

$$\mathbf{I} = \mathbf{X} \cdot \boldsymbol{\omega},$$

where $\mathbf{I}$ denotes the array of elements I_t and $\mathbf{X}$ is the matrix representing the integration operator:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

$\boldsymbol{\omega}$ is obtained by minimizing the following cost function:

$$F(\boldsymbol{\omega}) = \|\mathbf{I} - \mathbf{X} \cdot \boldsymbol{\omega}\|^2 + \alpha^2 \|\boldsymbol{\Gamma} \cdot \boldsymbol{\omega}\|^2 . \quad (3)$$

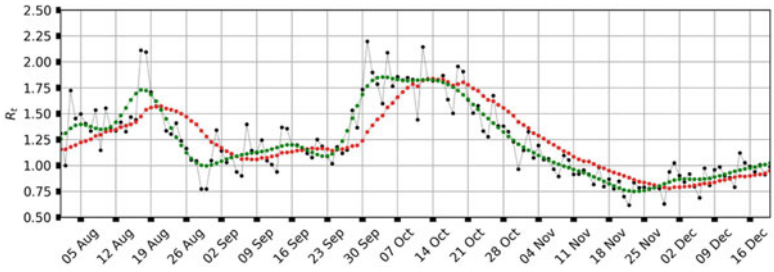


Fig. 13 In black R_t calculated on raw infected cases, in red the average R_t calculated over a window of 14 days; in green $\bar{R}_t$ calculated on the signal smoothed with Tikhonov regularization; each series is by symptoms onset date, data-set by ISS

Hence the derivative $\boldsymbol{\omega}$ is fitted using a Ridge regression with a generalized Tikhonov regularization factor:

- $\boldsymbol{\Gamma}$: the Tikhonov regularization matrix, chosen to be the second derivative operator;
- α: the regularization factor.

The regularization factor penalizes the spikes in the second derivative, forcing the derivative to be a smoothed signal. Once the derivative is fitted, the original signal is reconstructed by applying again the integral matrix to the differentiated smoothed signal; denoting the smoothed signal by $\bar{\mathbf{I}}$:

$$\bar{\mathbf{I}} = \mathbf{X} \cdot \boldsymbol{\omega}.$$

The parameter α can be obtained by searching the maximal smoothness constrained to the desired degree of information still available in the signal. It can be shown empirically that $\alpha = 100$ represents a reasonable trade-off.

Once we have $\bar{I}_t$ it can be fed into Eq. (2) to obtain the reproduction number computed on the smoothed signal, which we denote by $\bar{R}_t$.

Figure 13 shows in green the result of calculating $\bar{R}_t$ on the signal smoothed by minimizing Eq. (3). It is clearly visible that the green line anticipates the red one: maximizing the cross correlation of the original noisy R_t wrt $\bar{R}_t$ confirms a measure of the delay of 0 days.

5.4.2 Do Not Wait for the Symptoms Onset Date

Delays are not as important in literature, where we usually look at historical data, as they are in policy making, where we do need near real-time data.

Figure 14 shows the zoom of the series in Fig. 13 to the "present" days (which is Jul 30 at the time of writing). The last available "consolidated" count of new infected cases dates back to Jul 15, as the full process of data collection must be

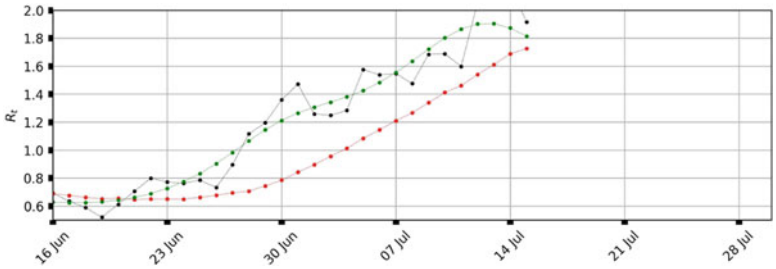


Fig. 14 Zoom of Fig. 13 to the most recent data available and consolidated, data-set by ISS

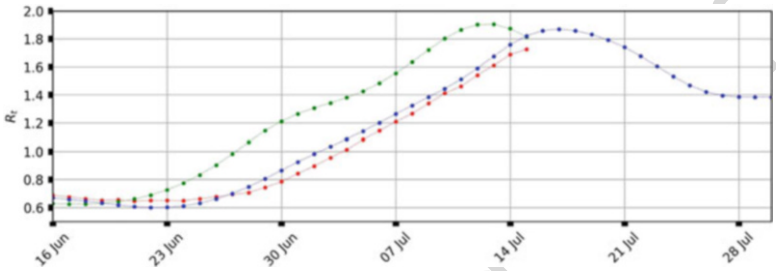


Fig. 15 In red the average R_t calculated over a window of 14 days on the symptoms onset distribution, data-set by ISS; in green $\bar{R}_t$ calculated on the symptoms onset distribution smoothed with Tikhonov regularization, data-set by ISS; in blue $\hat{\bar{R}}_t$ calculated on the notification date distribution smoothed, data-set by Protezione Civile

completed if we want to know the symptoms onset date. This problem, known as right censoring, is true for every country, with delays which vary depending on the particular data collection process. Moreover, it is well known in Italy that the collection process greatly depends on the pressure that the epidemic is producing on the Health System.

Instead of using the distribution of new infected cases by symptoms onset date, we propose to adopt the smoothed distribution by notification date as the input for Eq. (2), to obtain $\hat{\bar{R}}_t$. The difference is that as soon as a case is detected, it is notified. The advantage that the series is consolidated by definition, without the need of past revisions, comes with the following drawbacks:

1. there is a certain amount of delay from the symptoms onset date to the notification date;
2. the series accounts for more noise, as it makes no distinctions between symptomatic cases and asymptomatic cases.

Figure 15 shows the comparison of three R_t calculations: average R_t calculated on a 14-day window (in red), $\bar{R}_t$ calculated on smoothed cases by symptoms onset date (in green) and $\hat{\bar{R}}_t$ calculated on smoothed cases by notification date (in blue). The blue line exhibits a delay wrt the green one, but it is still anticipating the red line.

Maximizing the cross correlations provides the following measures of the relative delays:

- $\bar{R}_t$ anticipates $\hat{\bar{R}}_t$ by 8 days, but the last available value of $\bar{R}_t$ dates back 15 days prior to the present;
- $\hat{\bar{R}}_t$ anticipates R_t (calculated on a 14-day window) by 6 days.

Hence, we can conclude that, thanks to the smoothing procedure without delays (via Tikhonov regularization), we can replace the distribution of new cases by symptoms onset date with the distribution by notification date, obtaining the following advantages:

1. earn 6 days of anticipation with respect to the averaged R_t;
2. being able to compute the reproduction number up to the present, without having to wait for varying consolidation times in the data collection processes.

5.4.3 Residuals

As the original raw series are noisy and uncertain, we want a method to extract the noise and use it to calculate confidence intervals on the estimated R_t, in a way such that confidence intervals can directly reflect the uncertainty in the effective measuring process. This is much more relevant as we plan to estimate the reproduction number on the series of new infected cases by notification date, which includes both symptomatic and asymptomatic cases, with the latter exhibiting high noise.

The noise can be measured by the relative residuals of the signal with respect to its smoothed version, $\epsilon_t = (I_t - \bar{I}_t)/\bar{I}_t$.

Figure 16 shows that the distribution of ϵ_t calculated on the series of new infected cases by notification date **is unbalanced**.

It turns out that the unbalancing is directly related to the weekly seasonality which affects the series (the seasonality can be seen in Fig. 11 or Fig. 12). The reason is that the smoothing obtained by Eq. (3) is not able to capture the seasonality.

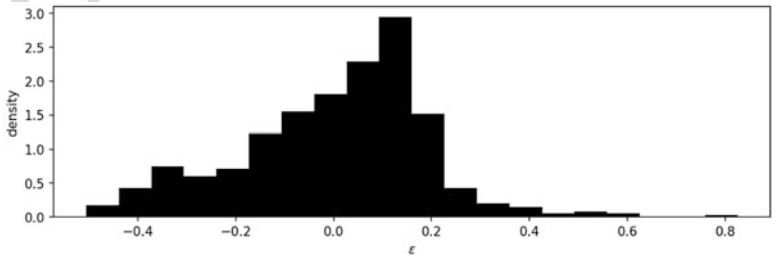


Fig. 16 Distribution of $\epsilon_t = (I_t - \bar{I}_t)/\bar{I}_t$ calculated on the series of new infected cases by notification date, data-set by Protezione Civile

5.4.4 Deseasoning via Singular Value Decomposition

Standard techniques to deal with seasonality, like SARIMA (Seasonal Autoregressive Integrated Moving Average), rely on moving averages.

To avoid the delays introduced by moving averages, we instead adopt Regularized Singular Value Decomposition (RSVD) proposed by Lin, Huang and Mcelroy in [28]. RSVD allows to detect the seasonal component of the signal by casting the signal vector into a matrix whose columns are the seasons and the rows are the repetitive periods of a complete series of seasons. Singular Value Decomposition is then applied to the matrix so that singular values represent the seasonal component of the signal. Each seasonal component is regularized via Tikhonov regularization, following the hypothesis that each seasonal component must change smoothly, period after period. The Tikhonov regularization parameter is fitted via "leave one out cross validation."

The advantage of this method with respect to the SARIMA approach is that *we do not need to take moving averages*, and we do not need to tune any meta-parameter of the model.

The python porting of the original R code is available in the supplementary material at https://github.com/tomorrowdata/COVID-19, within the library `covid19_pytoolbox`. The following features have been added to the original work:

- take the logarithm of the seasonal series, to remove exponential trends;
- differentiate the signal to a desired degree, to remove non-stationary trends in the original data, with an augmented Dickey–Fuller (ADF) test to check if any non-stationary component is present;
- apply Tikhonov regularization to the deseasoned signal to obtain the trend.

Denoting by $\tilde{I}_t$ the trend of the raw signal I_t after removing the seasonality, we obtain the following decomposition of the original series:

$$I_t = \tilde{I}_t + S_t + \tilde{E}_t \tag{4}$$

where S_t is the seasonal component and $\tilde{E}_t$ is the residual after deseasoning.

Figure 17 shows the result of applying RSVD to the series of new infected cases by notification date. RSVD has been applied to the logarithm of the second difference of the original series, with the ADF test confirming the removal of any non-stationary component. The smoothness of the seasonal components can be noted clearly.

5.4.5 Residuals of the Deseasoned Series

Now that we have removed the seasonality, we can look at the distribution of the residuals again. Figure 18 shows the distribution of the relative residuals after

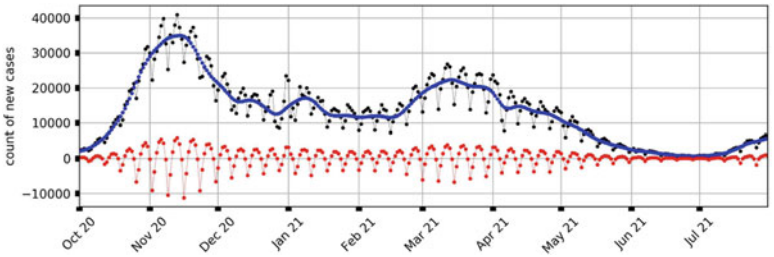


Fig. 17 Raw series I_t of new infected cases by notification date (in black), its trend $\tilde{I}_t$ after removing the seasonal component (in blue), the seasonal components S_t (in red); data-set Protezione Civile

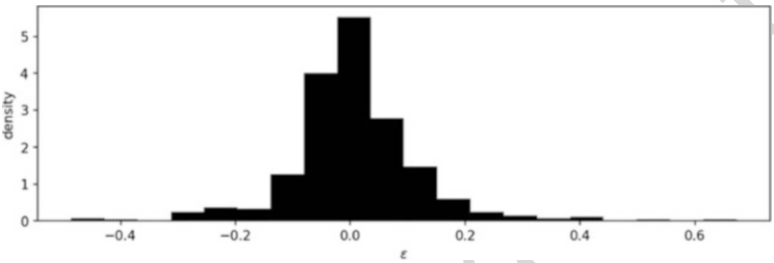


Fig. 18 Distribution of $\tilde{\epsilon}_t = \tilde{E}_t / \tilde{I}_t$ calculated on the series of new infected cases by notification date, data-set by Protezione Civile

removing the seasonal component: removing the seasonality produced a much more balanced, almost gaussian, distribution, if compared to Fig. 16.

5.4.6 Putting it All Together with Markov Chain Monte Carlo

We start from the series I_t of new infected cases by notification date, as explained in Sect. 5.4.2. We then apply RSVD to obtain the deseasoned smoothed trend $\tilde{I}_t$ of the series and the respective relative residuals $\tilde{\epsilon}_t = \tilde{E}_t / \tilde{I}_t$, as explained in Sect. 5.4.4.

With those ingredients, we can setup Markov chain Monte Carlo simulations to sample multiple chains of R_t values, as follows, denoting by $C(\cdot)$ the chains obtained via sampling:

1. $C(R_t)$ chains are sampled from a prior normal distribution, with $\mu = 1.3$ and $\sigma = 10$; a Gaussian process could be used instead, but it is less computationally efficient; the length of the chains is the same as the length of I_t;
2. $C(\tilde{\epsilon}_t)$ chains are sampled from a prior normal distribution, with $\mu = \sum_{s=t-7}^{t} \tilde{\epsilon}_s / 7$ and $\sigma = \sqrt{\sum_{s=t-7}^{t} (\tilde{\epsilon}_s - \mu)^2 / 7}$; the length of the chains is the same as the length of I_t;

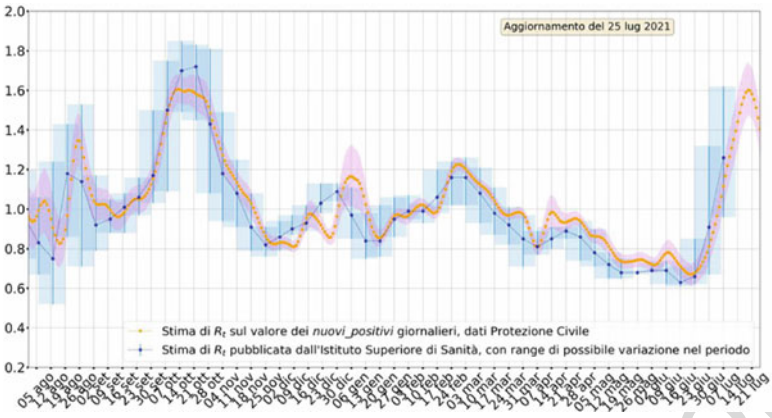


Fig. 19 In blue, the R_t values as reported by the Istituto Superiore di Sanità and in red the anticipated calculation published regularly, from the end of November 2020, at https://mondoeconomico.eu by Stefano Terna

3. $C(\tilde{I}_t)$, the chains of new cases with random noise, are obtained as $\tilde{I}_t + \tilde{I}_t \cdot C(\tilde{\epsilon}_t)$;
   - **note**: this is where the original noise of the series is transferred to the simulation, so that the confidence interval will account for uncertainties in the original series;
4. the estimated count T_t of new cases in each day of the chain is calculated from Eq. (2) as $C(T_t) = C(R_t) \cdot \sum_{s=1}^{t} w_s C(\tilde{I}_{t-s})$;
5. finally, a posterior Poisson distribution is tested via Monte Carlo, between the estimated cases, T_t, and the expected ones, $\tilde{I}_t$.

We sample 4 chains with 1000 iterations each discarded for tuning, and 500 iterations each kept for sampling. The final data-set contains 2000 samples from which day by day statistics, like the confidence interval, can be calculated.

Figure 19 shows the result, where the confidence interval (in violet) succeeds in representing periods of higher uncertainty in the data.

5.5 *Second Wave, New Infections from Outside, Introducing 20 Days in Advance the New Specific Containment Measures*

The counterfactual situation described in this section—inspired by Sect. 5.4—is related to the start and end dates of the actions of containment, both occurring 20 days in advance, with a natural barrier set on October 5th, 2020. Before that date, no one could plan to start new control measures.

As in the last two sections, we have 1407 cases of epidemics alive at the critical dates of June 1st and September 20th, after a Summer interval characterized by a quiet phase. Considering December 15th and September 20th situations, the second

wave epidemics are 769, again decreasing because the anticipated actions have eliminated some other cases. The December mean of the infected agents is 294.2, still higher than the actual value (≈200.0). We always overestimate the mean of the epidemic effects, being the surviving epidemics the larger ones.

Comparing Fig. 10d and c the difference is not evident; instead, the italic figures, and most of all, the red bold ones—in Table 7—report clearly the comparative advantage of this counterfactual experiment with respect to the values of Table 6.

5.6 Second Wave, New Infections from Outside, with a Unique Intervention Measure: Stopping Fragile People for 60 Days

The second counterfactual experiment is based on an immediate stop to the circulation of fragile persons and specifically of fragile workers, plus isolating nursing homes and hospitals. Schools are always open in this experiment. The decision is activated on October 5th, 2020, when the second wave was becoming evident. In [29] we have important consideration suggesting the importance of taking into account fragility in a long-term fighting perspective against this kind of epidemics.

As in the last three sections, we have 1407 cases of epidemics alive at the critical dates of June 1st and September 20th, after a Summer interval characterized by a quiet phase. Considering December 15th and September 20th situations, the second wave epidemics are 886, lightly above the values of Sects. 5.3 and 5.5, but without locking the economy and the society as a whole. The December mean of the infected agents is 326.3, higher than the actual value (≈200.0) for the explained overestimation bias.

Comparing Fig. 10e and c the difference is not evident; instead, the italic figures, and most of all, the violet bold ones—in Table 8—signal the close proximity of the effects of this counterfactual experiment with those of Table 6.

5.7 To Recap

Table 9 reports the different cases synthetically and, most of all allows an easy comparative interpretation of the actual and counterfactual situations.

6 Economic Analysis of the of Interventions

The pandemic has an impact on the general economy. First, we take into account the additional health expenditure, which in Piedmont has risen from € 8880 million

Table 7 Second wave, new infections from outside, with new specific measure anticipation of −20 days

t18.1	(1000)	Jun 1, 20		Sep 20, 20		Dec 15, 20		Feb 1, 21		May 1, 21		Dec 15, 20 to end		
t18.2	Cum. v.	Sym.	All	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	Totalinf.	Days
t18.3	Count	1407.0	1407.0	1407.0	1407.0	769.0	769.0	637.0	637.0	471.0	471.0	769.0	769.0	769.0
t18.4	Mean	35.6	72.7	40.0	84.1	**112.2**	**294.2**	*172.0*	*467.9*	*276.5*	*748.6*	248.9	663.4	499.3
t18.5	Std	14.1	42.6	16.7	52.8	66.8	188.4	91.5	251.3	112.9	286.9	158.0	417.5	124.1

Table 8 Second wave, new infections from outs., stop fragile people. 60 days from Oct. 5

t21.1	(1000)	Jun 1, 20		Sep 20, 20		Dec 15, 20		Feb 1, 21		May 1, 21		Dec 15, 20 to end		
t21.2	Cum. v.	Sym.	All	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	Totalinf.	Sympt.	Totalinf.	Days
t21.3	Count	1407.0	1407.0	1407.0	1407.0	886.0	886.0	761.0	761.0	637.0	637.0	886.0	886.0	886.0
t21.4	Mean	35.6	72.7	40.0	84.1	**128.1**	**326.3**	*211.0*	*555.1*	*323.3*	*862.1*	301.1	792.3	515.5
t21.5	Std	14.1	42.6	16.7	52.8	89.6	234.2	118.1	306.7	126.4	315.9	170.7	450.2	116.9

Table 9 Report of the key results, with count, mean, and std

Scenarios		Dec 15, 20		Dec 15, 20 to end			t24.1 t24.2
		Sympt.	Totalinf.	Sympt.	Totalinf.	Days	
No containment	Count	140.0	140.0	140.0	140.0	140.0	t24.3
In spontaneous	Mean	248.4	648.7	701.1	1757.9	594.2	t24.4
Second wave	Std	167.4	424.3	246.4	599.7	118.9	t24.5
No containment	Count	1044.0	1044.0	1044.0	1044.0	1044.0	t24.6
In forced	Mean	180.4	462.1	726.6	1810.9	620.9	t24.7
Second wave	Std	134.6	354.6	221.9	544.0	110.8	t24.8
Basic containment	Count	874.0	874.0	874.0	874.0	874.0	t24.9
In forced	Mean	*130.0*	*340.6*	252.7	666.4	494.1	t24.10
Second wave	Std	83.9	232.6	156.8	416.4	122.7	t24.11
−20 days cont.	Count	769.0	769.0	769.0	769.0	769.0	t24.12
In forced	Mean	**112.2**	**294.2**	248.9	663.4	499.3	t24.13
Second wave	Std	66.8	188.4	158.0	417.5	124.1	t24.14
Frag. subj. & workers control	Count	886.0	886.0	886.0	886.0	886.0	t24.15
In forced	Mean	**128.1**	**326.3**	301.1	792.3	515.5	t24.16
Second wave	Std	89.6	234.2	170.7	450.2	116.9	t24.17

to €9200 million, with an increase in pressure on GDP of 0.2%. It is an increment that cannot be generalized. In other regions and States, health expenditure has even decreased, due to the lower demand for diagnostic and treatment services, precisely because of the pandemic and the precautionary reduced access to health services. Apart from the additional health expenditure, the main impact to be considered is the loss in production induced by the contagion containment measures, i.e., the so-called lockdown of the economy and the associated mobility bans.

The impact assessment of production stoppages and mobility bans can be measured by applying an Input-Output model. The main quality of Input-Output models is the possibility of determining the total effect of changes in output in all sectors of the economy due to a unit change in final demand in a given sector. This is achieved by applying a matrix of multipliers, i.e., Leontief's inverse matrix, to a sectoral vector of demand changes. The inverse matrix makes it possible to calculate the sum of the direct impact of the stopped productions, sector by sector, and the indirect impact, due to the infinite feedback on the purchases of the affected sectors from the first drop in demand received. However, the standard representation is not complete. The literature tends to extend these effects to consider the feedback not only by the purchases of the impacted sectors but also by the drop in the final demand of households affected by the unexpected change in income through their marginal propensity to consume. This third effect is the so-called induced impact.

The matrix of direct, indirect, and induced impacts of the Piedmont economy in Table 10 has been originally estimated by one of the authors.

As we can see, the economic effects of lockdowns can be very different depending on whether they selectively affect one sector (normally the sectors most

Table 10 Multipliers of direct, indirect, and induced impact, and overall impact as well added value (GDP) multipliers per 1 euro of final demand change, related to each of the five sectors on the rows

	Final/total demand	Direct impact	Indirect impact	Total Induced impact	Production multiplier	Added value multiplier
Agriculture	0.53	1.40	0.50	1.30	3.20	1.40
Manufacturing	0.33	1.80	1.20	1.60	4.50	1.60
Construction	0.38	1.70	0.90	1.60	3.20	1.60
Distribution	0.53	1.40	0.50	1.30	3.10	1.60
Services	0.50	1.50	0.50	1.40	3.40	1.50

affected are the last two, distribution and services), or whether all sectors are affected. The manufacturing sector, which is strongly linked with other sectors, has a total, direct, indirect, and induced multiplication coefficient of 4.5 times the initial reduction in final demand. Therefore, to calculate the impacts, we started from three different assumptions, or scenarios, which we have called A, B, and C.

[A] The restrictions affected all economic activities that could be stopped, safeguarding only those businesses that were essential. This meant stopping approximately half of the regional production system. Schooling was only permitted with distance learning. This case occurred in the period from 9 March 2020 to the end of April 2020.

[B] Only businesses in sectors whose activities were rated with a high risk of contagion were stopped: these activities included non-food retail trade, the tourism restaurant and hotel sector, the sport, recreation and entertainment sector, the cultural sector, and, of course, the whole education system, that was served by distance learning. The transportation sector was legally active but still impacted by an almost obligatory drop in demand. This case actually occurred at different times during 2020 and 2021 and significantly from October 2020 until spring 2021, with a break of a few weeks during the winter.

[C] Purely theoretical and not put in place, it was considered to stop only the fragile workers, leaving intact the education and all the activities stopped in case B. In this case, fragile workers are estimated to be 14% of the total, based on a national projection of the total number of 5.6 million fragile people under 65 in Italy. To make the calculation of the impact realistic, we assume that all fragile workers received sickness compensation equal to the lost wage that impacted on the overall tax loss, increasing it; we also assume that half of the production of fragile workers could still be produced with overtime or temporary work by other workers.

The results of the simulations are reported in Table 11 and are expressed in points per thousand of Piedmont's GDP.

In Table 11, each day of closure of productive activities (leaving only the essential ones and schools closed, i.e., open for distance learning) to counter the

Table 11 Economic impacts of the pandemic with three hypotheses of non-pharmaceutical containment measures applied; values are expressed in GDP points/1000

	Scenario A	Scenario B	Scenario C	t29.1
Daily impacts				
Total production	−4.86	−1.23	−0.69	t29.2
Added value	−2.12	−0.55	−0.30	t29.3
Taxes	−0.91	−0.24	−0.35	t29.4
Monthly impacts				
Total production	−145.7	−36.9	−20.7	t29.5
Added value	−63.7	−16.6	−9.1	t29.6
Taxes	−27.4	−7.1	−10.6	t29.7

contagion and allow access to hospitals produces a loss of income (added value) equal to 2.1 per thousand of GDP and a worsening of the fiscal budget by 0.9 per thousand of GDP. One month of total closures, therefore, would cost an income loss of 6.4% and a worsening of the fiscal balance of 2.7% (Scenario A, actually implemented in Italy from 9 March 2020 to April).

Conversely, limiting closures to only distribution activities (non-food), as well as to school (open as distance learning), sports, culture and leisure, tourism, and restaurant services (as in the light lockdown established in October 2020 and subsequent months, Scenario B) would have produced a daily income loss of 0.55 per thousand of GDP (equivalent to 1.6% per month) and a fiscal loss of 0.24 per thousand per day and 0.7% of GDP per month.

The solution of protecting at home (and paying) only fragile workers, leaving all schools and productive activities open, would reduce the loss of income to 0.3 per thousand per day (0.91% per month). Although this solution (Scenario C, never actually implemented) is more convenient concerning the overall income loss, even 1/7 of that of scenario A and 1/2 of that of scenario B, it costs slightly more in fiscal terms than scenario B (−0.35 per thousand per day, instead of −0.24). However, it would seem preferable because it is the only option of the three that would allow the regular operation of the schools. According to the reliable Invalsi tests performed in 2021, the percentage of pupils in Italian schools who have not reached the minimum learning standards has increased by 10 percentage points based on the total number of pupils. If we were to put this loss of human capital on an economic balance sheet, we would have to consider the full cost of an additional year of schooling for 10% of the school population, both in terms of the cost of additional education plus the income lost for a 1-year delay in subsequent employment. A raw estimate of this cost would appear to be 1.3% of GDP, of which 0.58% for the additional cost of education and 0.75% for the income lost by postponing entry into employment by 1-year.

Following Table 12, in a C scenario, the cost of pandemic restrictions, for 3 months (hypothesis), would be 0.2% of annual GDP for increased health expenditures +2.7% of direct, indirect, and induced value-added (GDP) losses, plus 3.19% of GDP of public budget deterioration, while there would be no human capital losses. The total losses in scenario C would be 6.1% of annual GDP for three full

Table 12 Total losses simulating the three scenarios A, B, C, from activity and mobility restrictions, in GDP points/1000

	Scenario A three months	Scenario B three months	Scenario C three months
More health expenditure	−2.0	−2.0	−2.0
Added value or GDP loss	−191.0	−49.8	−27.2
Tax loss	−82.1	−21.4	−31.9
Human capital loss	−13.4	−13.4	0.0
Total loss (GDP/1000)	−288.6	−86.6	−61.1

months of restrictions. In scenarios B and A the total loss would have been much higher and specifically 8.6% and 28.8% of the pre-Covid GDP, respectively. It is also worth noting the distribution of losses by row. In scenario C, the losses in value-added, and thus the recession damage to the economy to be recovered, would be minimal, and the losses due to insufficient human capital formation would be zero. Nevertheless, the policies adopted have preferred the adoption of scenarios A and B.

7 Planning Vaccination Campaigns

7.1 Some Notes on Vaccines

Vaccines are biological products made from killed or attenuated microorganisms, from viruses or from some of their components (antigens), or from substances they produce made safe by chemical (e.g., formaldehyde) or heat treatment, while maintaining their immunogenic properties (https://www.who.int/vaccines); today, vaccines can be composed of proteins obtained by recombinant DNA techniques using genetic engineering approaches.

They usually contain, in addition to the antigenic fraction, sterile water (or a saline-based physiological solution), adjuvants, preservatives, and stabilizers. Adjuvants are included in the vaccine in order to enhance the immune system response; preservatives are added to prevent contamination of the preparate by bacteria; stabilizers are introduced to increase the shelf life of the product and to maintain the properties of the vaccine during storage.

How Vaccines Work: A Step Back in the Eighteenth Century

Although early forms of empirical immunization appear to have been present in different cultures (India and China; [30]), the creation of the first vaccine (for smallpox immunization) dates back to 1798 by Edward Jenner, an English physician. Jenner had noticed that milkmaids who became infected with cowpox (Vaccinia Virus), a virus that causes similar symptoms to human smallpox (Variola Virus or Smallpox virus) but not fatal, did not subsequently develop the disease [31].

This suggested that Cowpox inoculation could protect against Smallpox. Jenner decided to test his theory by inoculating an eight-year-old boy, the son of his gardener (sic!), James Phipps, with material taken from the cowpox lesions of a local milkmaid. As expected, James developed few local lesions and a modest fever. Two months later Jenner inoculated James with variolous matter from a case of human smallpox, without a sensible effect was produced: Jenner had proved that the boy had been immunized. By definition, all subsequent immunizations would be called vaccinations as in 1881 Louis Pasteur proposed it as a general term for the new protective inoculations, in honor of Jenner.

Once inoculated, vaccines (all of them), mimicking the first contact between man and pathogen, are able to stimulate an immunological response (humoral and cellular) as if this occurred through a natural contagion, although not leading to disease and without giving the associated complications. The rationale behind this phenomenon is immunological memory: the body/immune system that has already experienced a pathogenic microorganism, treasures the experience by responding rapidly to the same microorganism (the absence of immunological memory is the reason why Covid-19 emerged as a problem for humans). For some vaccines it is necessary to make recalls at a distance of time. Normally our body reacts to an unwanted host, but it can take up to two weeks to produce a sufficient amount of antibodies versus the pathogen. In the absence of vaccination, in this interval of time a pathogen can create damage to the body and even lead to death.

Types of Vaccines

A long way has been covered since 1798, and technologies have steadily improved to arrive at hi-tech vaccines such as those we are using today to fight Covid. The types of vaccine that exist today are:

- live attenuated vaccines (e.g., measles and tuberculosis): these are produced from infectious agents that have been rendered non-pathogenic;
- inactivated vaccines (e.g., poxvirus): these are produced using infectious agents killed by heat or chemicals;
- purified antigen vaccines (e.g., anti-meningococcal): these are produced by purifying specific components (bacterial or viral);
- anatoxin vaccines (e.g., tetanus): these are produced using molecules from the infectious agent, which are not capable of causing the disease on their own, but which can stimulate/activate the immune defenses of the vaccinee;
- recombinant protein vaccines (e.g., hepatitis B): these are produced using recombinant DNA technology, which involves inserting genetic material coding for the antigen (a protein/peptide) into microorganisms capable of producing the antigen specifically, allowing it to be purified;
- recombinant mRNA vaccines (e.g., Pfizer/BioNTec and Moderna): these are produced using an mRNA coding for a target gene encapsulated nanoparticle made with lipid bilayers; this information is able to drive the synthesis of an antigenic protein, through the cell machinery, and the triggering of the immunitary system;

- recombinant viral vector vaccines (e.g., Astrazeneca/Oxford): these are produced using an DNA coding for a target gene carried within a defective adenovirus (from human or from chimpanzee), able to vehicle the gene within the cell nucleus. This information is able to drive the synthesis of an antigenic protein, through the cell machinery, and the triggering of the immunitary system.

The latter types of vaccine (mRNA and Adenoviral vector-based) were today adopted mainly because they can be manufactured very quickly, and being their production process highly standardized. As a matter of fact, they are the fastest way to create a vaccine in the middle of a pandemic.

mRNA and Adenovirus-Based Vaccines

Let us take a step back. The CoV-SARS-2019 virus has on its surface a protein called Spike (S-protein), that it uses to enter a human cell via binding to the ACE2 receptor (Fig. 3). The S-protein has therefore been chosen as the specific target to produce a vaccine since it is exposed in large quantities on the surface of the virus.

mRNA-based and adenovirus-based vaccine for Covid target the S-protein through the production of a RNA messengers (mRNAs), the classical molecule that routinely instructs all the cells what to build. Once the S-protein is produced within the body and presented the immune system, it is considered an antigen, and the body starts producing antibodies against it. The same thing can be done by using a pre-made protein and injecting it, but its production, testing and approval is longer (years to decades) and more expensive.

In a mRNA-based vaccine (e.g., Pfizer/Moderna) the mRNA coding for the Cov-Sars-2019 spike (S-protein) is encapsulated in lipid nanoparticles. This preparate is then injected (usually in the deltoid muscle). After that, the nanoparticles fuse with the cell membranes and mRNA is released into the cell cytoplasm, without entering in the nucleus nor getting incorporated into the genomic DNA. In an adenovirus-based vaccine (e.g., AstraZeneca) the gene coding for the spike protein is inglobated as DNA in a defective Chimpanzee adenovirus which is not able to proliferate, alone. This virus once injected latch on the host cell and released DNA (carrying the spike protein gene) in the cell cytoplasm. DNA then migrates to the nucleus where it is transcribed into mRNA, which will migrate to the cytoplasm.

In both the vaccines, at this particular stage the mRNA coding for the Spike uses the cellular machinery (e.g., ribosomes) for being translated into protein, imitating virus-infection-like humoral immunity and cellular immunity [32]. Both mRNA-based and adenovirus-based vaccines are able to increase the host's anti-virus effects by increasing T cells' antigen reactiveness [33]. Normally, are these white blood cells, as the first defenses of the body, to "detect" the presence of the pathogen and to organize a protection by generating specific antibodies to combat it through B-lymphocytes, as particular blood cells deputed to antibody production [34] These antibodies cover the virus and prevent it from attacking our body.

The immune system memory can be compared to a human's memory. Once it encounters an unwanted visitor, it will remember it and will be able to recognize it in the future. This process typically takes a few weeks for the body to produce the antibodies, but these cells will be there to guard the body for a long time.

7.2 *Planning a Vaccination Campaign Using Genetic Algorithms, with Non-pharmaceutical Containment Measures in Action*

We compare the effect of choosing the vaccination quotas via genetic algorithms (GAs) with two predetermined strategies. Our model considers three hypotheses: vaccinated people still spread the contagion; they do not spread the contagion; they do it in the 50% of the case. We show here only the results of the first case, the worst (as we write, the Delta variant is spreading, with vaccinated people transmitting the infection).

The parameters of the GAs side of the model are contained in a special file, as described in the Info sheet of the model; at https://terna.to.it/simul/SIsaR.html start the model and look at the Info paragraph named *Using Genetic Algorithms*.

Important dates:

- in the internal calendar of the model, day 373 is February 12th, 2021; it is the starting point of the vaccinations in Piedmont;
- the effectiveness of the initial vaccinations, 40 days later, starts on day 413 (March 22nd, 2021).

A technical detail: we simulate the vaccination campaigns with the GAs using the BehaviorSearch program, https://www.behaviorsearch.org, strictly related to NetLogo.

7.2.1 Vaccination Groups

We take into consideration seven groups, in order of decreasing fragility, also considering the exposure to contagions:

g1 Extra-fragile people with three components;

- due to intrinsic characteristics: people in living in nursing homes;
- due to risk exposure:
  - nursing homes operators;
  - healthcare operators;

g2 teachers;
g3 workers with medical fragility;
g4 regular workers;
g5 fragile people without special characteristics;
g6 regular people, not young, not worker, and not teacher;
g7 young people excluding special activity cases (a limited number in *g1*).

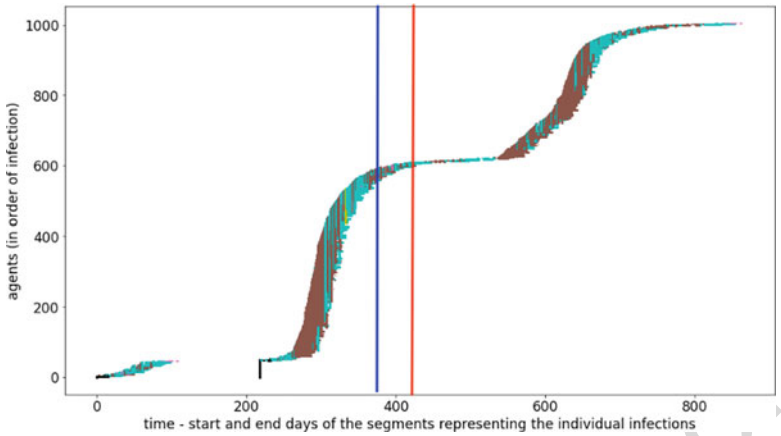


Fig. 20 Crucial dates: blue line for the starting point of the vaccination campaign and red line for the start of the effectiveness of the initial vaccinations; all the situations without vaccination

7.3 A Specific Realistic Case

The description of the vaccination effects on an outbreak is quite lengthy. Considering the collection at https://terna.to.it/simul/GAresultPresentation.pdf, we report here a unique case: the experiment I reported there, maintaining the reference to I in the titles of the figures. Considering the adoption of the government non-pharmaceutical measures, we search—in the batch of the 10,000 outbreaks of Sect. 5.3—for realizations of sequences similar to the actual events that occurred in Piedmont. As we see in Fig. 20 and the related Fig. 21, the artificial case that we adopt for the GAs exploration has the following critical characteristics:

(i) numbers of infected persons quite similar at `cp2` and at `cp3` in Fig. 9a; besides, numbers not too different from those of the same figure;
(ii) number of infected persons at `cp4` significantly greater than those at the previous checkpoint.

In Fig. 21, without vaccinations, we have the first wave in Spring 2020, a larger one in Fall 2020, a limited one between the end of 2020 and the beginning of 2021; then, a relatively quiet interval and successively, just while we write these notes, some restarting signals; finally, a fourth wave. Currently, it is in the future, relative to both the time of writing and the time when the calculations were completed (see NB at the end of Sect. 4.2). Very realistic with Piedmont's actual situation, the limited thickness of the *snake* of Fig. 20, when vaccinations start and when their effectiveness develops. The hole in the series identifies a period of quasi-extinct epidemics. Then it restarts with the arrival of infected persons from outside.

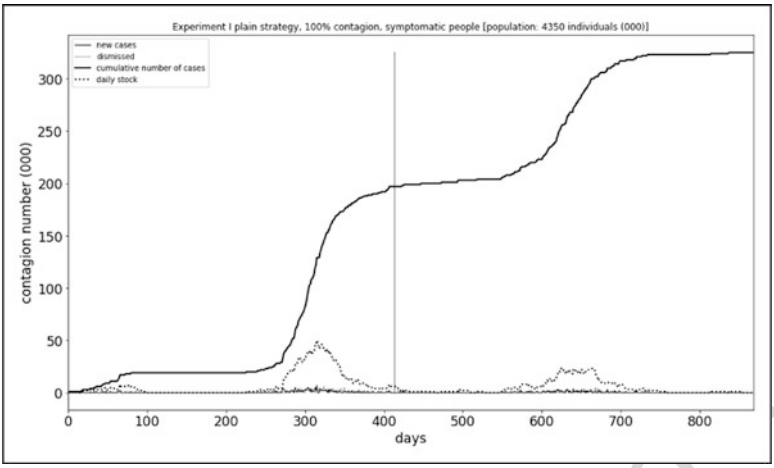


Fig. 21 Base symptomatic series; the vertical line at day 413 is not relevant here

Here and in the following sections, we analyze the count of symptomatic persons, being the goal of our simulated vaccination campaign exactly that of decreasing the number of symptomatic people, as deceased persons come from there.

7.4 Vaccination Quotas, Plain *Strategy*

The vaccination plans are related to the first dose; the second dose is supposed to be automatically scheduled, with an independent supply. The vaccinated person starts to benefit from immunity 40 days after the first dose.

Considering a *plain* option as that adopted in Table 13 with, in each day, the quantities of doses of the first column, we will primarily vaccinate the left column groups to move gradually to people of the other columns, as those on the left have already received the vaccine. The order is (*g1*) extra-fragile people, (*g2*) teachers, (*g3*) fragile workers, (*g4*) regular workers, (*g5*) fragile people, (*g6*) regular people, (*g7*) young people. In Table 14 we have numbers both of persons in each category at the beginning of this experiment (and in the following ones) and when the vaccination campaign starts.

Some of the coefficients in Table 13, and all the successive similar ones, are not used in two situations:

(i) when the persons of a group are fully vaccinated, the quotas in the rows below that day are not relevant;
(ii) when the people in the columns to the left of a given column completely absorb the available doses of vaccine on that day (the quotas in that column have unimportant values).

Table 13 From the day of the first column, considering the quantity of the second column (000), the vaccination of each group follows the quotas of the related columns

From day	Q. of vaccines (000)	*g1*	*g2*	*g3*	*g4*	*g5*	*g6*	*g7*
373	5	0.1	0.1	0.1	0.1	0.1	0.1	0.1
433	10	0.1	0.1	0.1	0.1	0.1	0.1	0.1
493	10	0.1	0.1	0.1	0.1	0.1	0.1	0.1
553	10	0.1	0.1	0.1	0.1	0.1	0.1	0.1
613	20	0.1	0.1	0.1	0.1	0.1	0.1	0.1
738	End							

Table 14 Susceptible persons at the beginning of the simulation and when the vaccination campaign starts, day 373, Feb. 12th, 2021

(000)	*g1*	*g2*	*g3*	*g4*	*g5*	*g6*	*g7*
Susc. at $t = 0$	133	84	240	1560	1179	254	900
Susc. when vacc. starts	124	81	162	1234	1032	245	891

We anticipate that the GAs procedure does not optimize the coefficients of cases (i) and (ii).

The series that we introduce hereafter are significant from day 413, March 22nd, when the initial vaccinations' effectiveness begins, after 40 days from initial vaccinations.

In Fig. 22a we have the effects of the vaccination plan as numbers of vaccinated persons by groups. In Fig. 22b we have the most important outcome: the no vaccination test-bed is that of Fig. 21. We note the waves after the vertical line—when vaccinations start to operate—are lower than in the test plot, but anyway, those further waves are there.

7.5 *Vaccination Quotas,* Wise *Strategy*

Considering now a *wise* option, as an attempt to mimic the actual (and complex) vaccine distribution in the region, we use the quotas of Table 15, with the exact mechanism of the previous section. We primarily vaccinate the left column groups to move gradually to other columns, but postponing group *g4* (regular workers), *g6* (regular people), and *g7* (young people). In Table 14 we have numbers both of persons in each category at the beginning of this experiment (and in the following ones) and when the vaccination campaign starts. The considerations sub (i) and (ii) in Sect. 7.4 apply also here.

In Fig. 22c we have the effects of the vaccination plan as numbers of vaccinated persons by groups. In Fig. 22d we have the experiment outcome: the no vaccination test-bed is that of Fig. 21. We note the waves after the vertical line—when

Table 15 From the day of the first column, considering the quantity of the second column (000), the vaccination of each group follows the quotas of the related columns

From day	Q. of vaccines (000)	*g1*	*g2*	*g3*	*g4*	*g5*	*g6*	*g7*
373	5	0.1	0.1	0.1	0.0	0.1	0.0	0.0
433	10	0.1	0.1	0.1	0.0	0.1	0.0	0.0
493	10	0.1	0.1	0.1	0.1	0.1	0.1	0.1
553	10	0.1	0.1	0.1	0.1	0.1	0.1	0.1
613	20	0.1	0.1	0.1	0.1	0.1	0.1	0.1
738	End							

vaccinations start to operate—are lower than in the test plot, but we have significant further waves in this case too.

7.6 GAs Quotas in the Experiment, with Vaccinated People Spreading the Infection

Finally, this whole section's objective is to use GAs to evolve populations of models by choosing "genetically" the parameters to decide daily vaccination. Initially, on a random basis and successively considering them as a genetic chromosome of each model, re-productively crossed with those of other models. The search is for the best fitness related to the goal of reducing the number of symptomatic persons. [35], also quoted at https://www.behaviorsearch.org, is a helpful introduction to the methodology; the sources of the GAs used here are at https://github.com/terna/GAs. The GAs action, determining the vaccination quotas, optimizes the behavior of a *deciding* meta-agent, in a sort of *inverse generative social science perspective* [36].

With the GAs option, we use the quotas of Table 16, with the exact mechanism of the previous section. The considerations sub (i) and (ii) in Sect. 7.4 also apply here. We underline that the GAs procedure does not optimize the coefficients of those two cases, because they do not affect the fitness related to the goal of minimizing the number of symptomatic subjects.

In Table 14 we have numbers both of persons in each category at the beginning of the experiment and when the vaccination campaign starts.

In Fig. 22e we have the effects of the vaccination plan as numbers of vaccinated persons by groups. The main attention of the GAs is initially related to the groups: *g4* (workers), *g1* (extra-fragile persons), *g3* (fragile workers), *g2* (teachers). Then *g5* (fragile people), finally *g6* (regular people), and *g7* (young people). The priority is for highly circulating persons (workers and teachers), then for fragile persons.

In Fig. 22f we have the crucial result of this experiment: the no vaccination test-bed is always that of Fig. 21. With GAs' choices, the waves after the vertical

Table 16 GAs best strategy with *vaccinated people still spreading the infection*: from the day of the first column, considering the quantity of the second column, the vaccination of each group follows the quotas of the related columns

From day	Q. of vaccines (000)	*g1*	*g2*	*g3*	*g4*	*g5*	*g6*	*g7*
373	5	0.01	0	0	0.79	0.18	0.38	0.19
433	10	0.94	0.06	0.32	0.54	0.19	0.83	0.5
493	10	0.97	0.97	0.74	0.79	0.2	0.14	0.52
553	10	0.98	0.83	0.02	0.39	0.99	0.04	0.48
613	20	0.52	0.01	0.83	0.6	1	0.27	0.9
738	End							

line—when vaccinations start to operate—disappear, and the whole outbreak is a lot shorter.

8 A New Model and Future Developments

Using SLAPP, https://terna.github.io/SLAPP/ a second model is under development, with a ratio of 1:100 to the Piedmont population, so 43,500 agents. It will contain the same items as the current one, plus transportation and aggregation places: happy hours, nightlife, sports, stadiums, discotheques, etc. We will also consider networks as family networks, professional networks, high-contact individual networks [37]. Finally, we will take into consideration the socioeconomic conditions of the individuals.

As seen, the S.I.s.a.R. model is a tool for comparative analyses, not for forecasting, mainly due to the enormous standard deviation values intrinsic to the problem.

The model is highly parametric, and more it will be, precisely in the comparative perspective. It also represents a small step in using artificial intelligence tools and the inverse generative perspective [8] in agent-based models.

9 Appendix 1—Parameter Values

We report here the values of parameters of Fig. 5, with their short names used in program scripts, in round brackets. Look at Sect. 2.2 for the definition. Day numbering is related to actual dates via Table 17. Day 1 is February 4th, 2020.

The values adopted in the experiments reported in this work are the following.

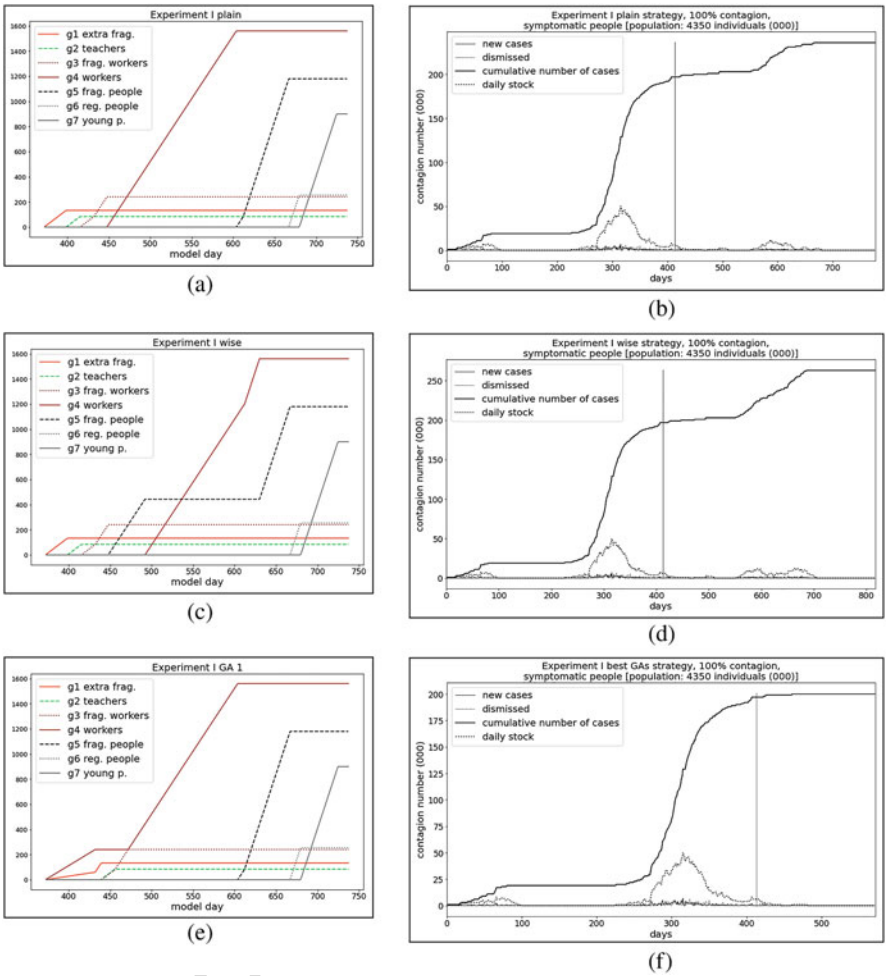


Fig. 22 Vaccination sequences and time series. (**a**) *Plain* vaccination sequence; on the *y* axis the number of vaccinated subjects of each group (if vaccination is complete, the line is horizontal). (**b**) *Plain* vaccination symptomatic series; the vertical line is at day 413, when the effectiveness of first vaccination starts. (**c**) *Wise* vaccination sequence; on the *y* axis the number of vaccinated subjects of each group (if vaccination is complete, the line is horizontal). (**d**) *Wise* vaccination symptomatic series; the vertical line is at day 413, when the effectiveness of first vaccination starts. (**e**) GA vaccination sequence; on the *y* axis the number of vaccinated subjects of each group (if vaccination is complete, the line is horizontal). (**f**) GAs vaccination symptomatic series; the vertical line is at day 413, when the effectiveness of first vaccination starts

(i) The values of *probabilityOfGettingInfection* (`prob`) are: 0.05 (starting phase); 0.02 at day 49 (adoption of non-pharmaceutical measures); 0.035 at day 149 (some relaxation in compliance); 0.02 at day 266 (again, compliance to rules).

Table 17 The days of the simulation and their equivalent dates in the calendar

Day	Date	Day	Date	Day	Date	Day	Date
25	28-2-2020	200	21-8-2020	375	12-2-2021	550	6-8-2021
50	24-3-2020	225	15-9-2020	400	9-3-2021	575	31-8-2021
75	18-4-2020	250	10-10-2020	425	3-4-2021	600	25-9-2021
100	13-5-2020	275	4-11-2020	450	28-4-2021	625	20-10-2021
125	7-6-2020	300	29-11-2020	475	23-5-2021	650	14-11-2021
150	2-7-2020	325	24-12-2020	500	17-6-2021	675	9-12-2021
175	27-7-2020	350	18-1-2021	525	12-7-2021	700	3-1-2022

(ii) The value of *D%* is −50 in all the runs.

(iii) *intrinsicSusceptibility* is set discussing Eq. (1) in Sect. 2.2.

(iv) The values of *%PeopleAnyTypeNotSymptomaticLeavingHome* (`%PeopleAny`) are: at (day) 20, 90; at 28, 80; at 31, 0; at 106, 80; at 110, 95; at112, 85; at 117, 95; at 121, 90; at 259, 90; at 266, 80; at 277, 50; at 302, 70; at 320, 90; at 325, 50; at 329, 80; at 332, 50; at 336, 80; at 337, 50; at 339, 80.

(v) The values of *%PeopleNotFragileNotSymptomaticLeavingHome* (`%PeopleNotFr`) are: at (day) 31, 80; at 35, 70; at 36, 65; at 38, 15; at 42, 25; at 84, 30; at 106, 0; at 302, 90; at 325, 50; at 332, 50; at 337, 50; at 339, 100; at 349, 90.

(vi) The values of *%openFactoriesWhenLimitationsOn* (`%Fac`) are: at (day) 38, value4 0; at 49, 20; at 84, 70; at 106, 100; at 266, 90; at 277, 70; at 302, 80; at 320, 90; at 325, 30; at 329, 90; at 332, 30; 336, 90; at 337, 30; at 339, 100.

(vii) *stopFragileWorkers* (`sFW`): by default, 0; in one of the experiments we used `sFW` with set to 1 (on) at day 245 and to 0 (off) at day 275.

(viii) The values of *activateSchools* (`aSch`) are: at (day) 1, on; at 17, off; at 225, on; at 325, off; at 339, on; the values of *%Students* (`%St`) are: at (day) 0, 100; at 277, 50; at 339, 50; at 350, 50 (repeated values are not relevant for the model, but for the use of the programmer-author).

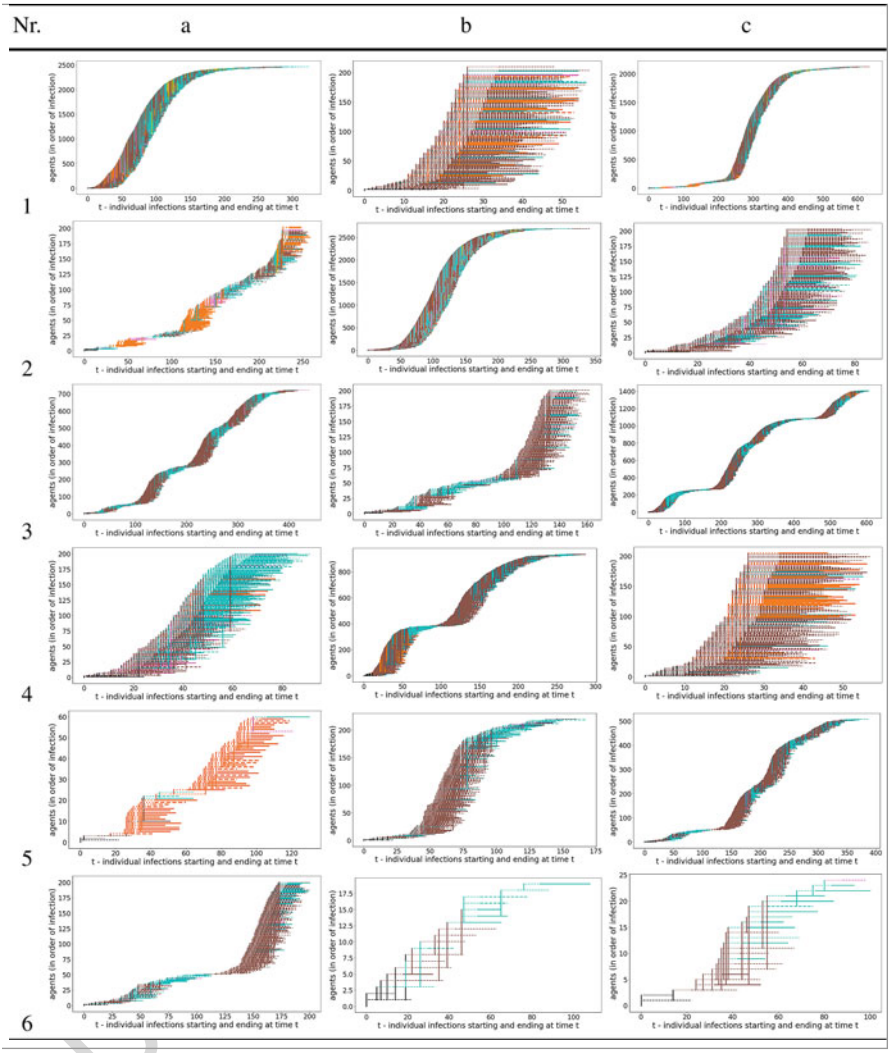
(ix) The value of *radiusOfInfection* (`radius`) is 0.2; in the model, space is missing of a scale, but forcing the area to be in the scale of a region as Piedmont, 0.2 is equivalent to 20 m; we have to better calibrate this measure with movements and probabilities; this is a critical step in future developments of the model.

(x) The values of *asymptomaticRegularInfected%* and *asymptomaticFragileInfected%* are 95 and 20.

10 Appendix 2—A Gallery of Contagion Sequences

The gallery of contagion sequences, reported in Table 18, shows the vast variety of situations generated by our agent-based simulations. What is significant is the variety of the situations.

Table 18 Gallery of sequences, symptomatic, and asymptomatic agents



t50.1

(1a) An outbreak without containment measures, with a unique wave, but very heavy: contagions are in nursing homes (orange), workplaces (brown), homes (cyan), hospitals (pink).

(1b) This is the previous epidemic without containment measures, considering the first 200 infections, with the main contribution of nursing homes (orange) and workplaces (brown).

(1c) Another outbreak, always without containment measures: nursing homes (orange) as a starter.

(2a) The (1c) epidemic, without containment measures, first 200 infections: nursing homes (orange) as a starter; around day 70, a unique contagion at home makes the epidemic continue.

(2b) Another case without containment measures showing the initial action of contagions in workplaces (brown) and homes (cyan).

(2c) Here we see the first 200 infections showing that the initial profound effects of contagions in workplaces (brown) and homes are due, in the beginning, to fragile persons, also asymptomatic,

(3a) An outbreak with containment measures, where we see another influential contribution of workplaces (brown) and homes (cyan) to the epidemic diffusion.

(3b) Here the first 200 infections: after day 100, we observe many significant cases of fragile workers diffusing the infection.

(3c) In this outbreak, with containment measures, the infections arise from workplaces (brown), nursing homes (orange), and homes (cyan), but also hospitals (pink).

(4a) Here we explore the first 200 infections of (3c): in the beginning, workplaces (brown), hospitals (pink), nursing homes (orange), and homes (cyan) are interweaving.

(4b) An outbreak with containment measures where the effect of the contagions in workplaces (brown), nursing homes (orange), and homes (cyan) is evident.

(4c) In the first 200 infections of (4b), workplaces (brown) and nursing homes (orange) are strictly interweaving.

(5a) An outbreak with containment measures where the effect of nursing homes (orange) is prevalent.

(5b) An outbreak with containment measures with a highly significant effect from workplaces (brown).

(5c) Stopping fragile workers at day 20 in the previous case, we obtain a beneficial effect, but home contagions (cyan) keep alive the pandemic, which explodes again in workplaces (brown).

(6a) Exploring the first 200 infections of the case (5c), we have evidence of the event around day 110 with the new phase due to a unique asymptomatic worker.

(6b) Finally, the same epidemic stopping fragile workers and any fragility at day 15 case and isolating nursing homes.

(6c) An outbreak with containment measures spontaneously stopping in a short period.

Acknowledgments Many thanks to Simone Landini, Nizar Mansour, Matteo Morini, Fabio Pammolli, Enrico Scalas, and Federico Tedeschi for their highly valuable discussions, insights, and critics. The usual disclaimer applies.

To run our model, we tremendously benefit from the use of HPC facilities provided by: Sarah de Nigris and Matteo Morini; http://tomorrowdata.io; the HP4AI (https://hpc4ai.unito.it/) and C3S centers at University of Torino.

References

1. A. Scala, A. Flori, A. Spelta, E. Brugnoli, M. Cinelli, W. Quattrociocchi, F. Pammolli, Scientific Reports **10**(1), 13764 (2020). https://doi.org/10.1038/s41598-020-70631-9
2. N. Bellomo, R. Bingham, M.A.J. Chaplain, G. Dosi, G. Forni, D.A. Knopoff, J. Lowengrub, R. Twarock, M.E. Virgillito, Mathematical Models and Methods in Applied Sciences **30**(08), 1591 (2020). https://doi.org/10.1142/S0218202520500323
3. H. Rahmandad, J. Sterman, Management Science **54**(5), 998 (2008)
4. F. Squazzoni, J.G. Polhill, B. Edmonds, P. Ahrweiler, P. Antosz, G. Scholz, E. Chappin, M. Borit, H. Verhagen, F. Giardini, N. Gilbert, Journal of Artificial Societies and Social Simulation **23**(2), 10 (2020). https://doi.org/10.18564/jasss.4298. http://jasss.soc.surrey.ac.uk/23/2/10.html
5. J.B. Bak-Coleman, M. Alfano, W. Barfuss, C.T. Bergstrom, M.A. Centeno, I.D. Couzin, J.F. Donges, M. Galesic, A.S. Gersick, J. Jacquet, A.B. Kao, R.E. Moran, P. Romanczuk, D.I. Rubenstein, K.J. Tombak, J.J. Van Bavel, E.U. Weber, Proceedings of the National Academy of Sciences **118**(27) (2021). https://doi.org/10.1073/pnas.2025764118. https://www.pnas.org/content/118/27/e2025764118
6. J. Epstein, Journal of Artificial Societies and Social Simulation **11**(4), 12 (2008). http://jasss.soc.surrey.ac.uk/11/4/12.html
7. R. Axtell, Why agents? On the varied motivations for agent computing in the social sciences. Tech. rep., Center on Social and Economic Dynamics Brookings Institution (2000)
8. T.M. Vu, C. Probst, J.M. Epstein, A. Brennan, M. Strong, R.C. Purshouse, Genetic and Evolutionary Computation Conference : [proceedings]. Genetic and Evolutionary Computation Conference **2019**, 1356 (2019). https://doi.org/10.1145/3321707.3321840. https://pubmed.ncbi.nlm.nih.gov/33083795
9. J.M. Epstein, Complexity **4**(5), 41 (1999). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.118.546&rep=rep1&type=pdf
10. G. Pescarmona, in *Recent research developments in biophysical chemistry*, ed. by C.A. Condat, A. Baruzzi (Trivandrum, 2002), pp. 1–22. http://flipper.diff.org/static/files/1517/The_life_context_2002.pdf
11. L. Gennero, M.A. Roos, P. D'Amelio, T. Denysenko, E. Morra, K. Sperber, V. Ceroni, M. Panzone, F. Lesca, E. De Vivo, A. Grimaldi, M.L. Gabetti, A. Ponzetto, G.P. Pescarmona, A. Pugliese, Cell Biochemistry and Function **28**(2), 142 (2010)
12. A. Vernone, P. Berchialla, G. Pescarmona, PloS one **8**(4), e60220 (2013). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0060220
13. A. Vernone, C. Ricca, D. Merlo, G. Pescarmona, F. Silvagno, Royal Society open science **6**(4), 181891 (2019). https://royalsocietypublishing.org/doi/pdf/10.1098/rsos.181891
14. F. Scialo, A. Daniele, F. Amato, L. Pastore, M.G. Matera, M. Cazzola, G. Castaldo, A. Bianco, Lung pp. 1–11 (2020). https://link.springer.com/article/10.1007/s00408-020-00408-4
15. S. Shenoy, Inflammation Research **69**(11), 1077 (2020). https://doi.org/10.1007/s00011-020-01389-z. https://doi.org/10.1007%2Fs00011-020-01389-z
16. F. Silvagno, A. Vernone, G.P. Pescarmona, Antioxidants **9**(7), 624 (2020). https://doi.org/10.3390/antiox9070624
17. M. Soy, G. Keser, P. Atagündüz, F. Tabak, I. Atagündüz, S. Kayhan, Clinical rheumatology **39**, 2085 (2020). https://link.springer.com/content/pdf/10.1007/s10067-020-05190-5.pdf
18. B. Hu, S. Huang, L. Yin, Journal of medical virology **93**(1), 250 (2021). https://onlinelibrary.wiley.com/doi/full/10.1002/jmv.26232
19. Y.d. Gao, M. Ding, X. Dong, J.j. Zhang, A. Kursat Azkur, D. Azkur, H. Gan, Y.l. Sun, W. Fu, W. Li, et al., Allergy **76**(2), 428 (2021). https://onlinelibrary.wiley.com/doi/full/10.1111/all.14657
20. A. Ghorbani, F. Lorig, B. de Bruin, P. Davidsson, F. Dignum, V. Dignum, M. van der Hurk, M. Jensen, C. Kammler, K. Kreulen, L.G. Ludescher, A. Melchior, R. Mellema, C. Păstrăv, L. Vanhée, H. Verhagen, Review of Artificial Societies and Social Simulation (2020). https://rofasss.org/2020/04/25/the-assocc-simulation-model/

21. R. Horton, Lancet (London, England) **396**(10255), 874 (2020). https://www.thelancet.com/action/showPdf?pii=S0140-6736%2820%2932000-6
22. U. Wilensky. Netlogo (1999). http://ccl.northwestern.edu/netlogo/
23. P. Steinmann, J.R. Wang, G.A. van Voorn, J.H. Kwakkel, Review of Artificial Societies and Social Simulation **17** (2020). URL https://rofasss.org/2020/04/17/deep-uncertainty/
24. L.M. Bettencourt, R.M. Ribeiro, PloS one **3**(5), e2185 (2008). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0002185
25. A. Cori, N.M. Ferguson, C. Fraser, S. Cauchemez, American Journal of Epidemiology **178**(9), 1505 (2013). https://doi.org/10.1093/aje/kwt133
26. R. Anderson, C. Donnelly, D. Hollingsworth, M. Keeling, C. Vegvari, R. Baggaley, R. Maddren, The Royal Society **2020** (2020). https://royalsociety.org/-/media/policy/projects/set-c/set-covid-19-R-estimates.pdf
27. D. Cereda, M. Tirani, F. Rovida, V. Demicheli, M. Ajelli, P. Poletti, F. Trentini, G. Guzzetta, V. Marziano, A. Barone, M. Magoni, S. Deandrea, G. Diurno, M. Lombardo, M. Faccini, A. Pan, R. Bruno, E. Pariani, G. Grasselli, A. Piatti, M. Gramegna, F. Baldanti, A. Melegaro, S. Merler. The early phase of the covid-19 outbreak in Lombardy, Italy (2020). https://arxiv.org/abs/2003.09320
28. W. Lin, J.Z. Huang, T. McElroy, Journal of Business & Economic Statistics **38**(3), 487 (2020). https://doi.org/10.1080/07350015.2018.1515081
29. N. Phillips, Nature **590**(7846), 382 (2021). https://www.nature.com/articles/d41586-021-00396-2
30. A. Boylston, Journal of the Royal Society of Medicine **105**(7), 309 (2012). https://doi.org/10.1258/jrsm.2012.12k044
31. E. Jenner, *An inquiry into the causes and effects of the variolae vaccinae, a disease discovered in some of the western counties of England, particularly Gloucestershire, and known by the name of the cow pox* (Springfield [Mass.] : Re-printed for Dr. Samuel Cooley, by Ashley & Brewer, 1802, 1800). http://resource.nlm.nih.gov/2559001R
32. M.A. Monslow, S. Elbashir, N.L. Sullivan, D.S. Thiriot, P. Ahl, J. Smith, E. Miller, J. Cook, S. Cosmi, E. Thoryk, et al., Vaccine **38**(36), 5793 (2020). https://www.sciencedirect.com/science/article/pii/S0264410X20308483
33. Y. Wang, Z. Zhang, J. Luo, X. Han, Y. Wei, X. Wei, Molecular Cancer **20**(1), 33 (2021). https://doi.org/10.1186/s12943-021-01311-z
34. W. Ratajczak, P. Niedźwiedzka-Rystwej, B. Tokarz-Deptuła, W. Deptuła, Central-European journal of immunology **43**(2), 194 (2018). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6102609/
35. J.H. Miller, Management Science **44**(6), 820 (1998). https://www.jstor.org/stable/pdf/2634650.pdf
36. T.M. Vu, C. Probst, J.M. Epstein, A. Brennan, M. Strong, R.C. Purshouse, in *Proceedings of the Genetic and Evolutionary Computation Conference* (2019), pp. 1356–1363. https://dl.acm.org/doi/abs/10.1145/3321707.3321840
37. G. Manzo, A. van de Rijt, Journal of Artificial Societies and Social Simulation **23**(4), 10 (2020). http://jasss.soc.surrey.ac.uk/23/4/10.html

AUTHOR QUERIES

AQ1. Please check the presentation of the authors' affiliation, and correct if necessary. **OK, all the affiliations are fine**

AQ2. Please check the sentence "The sequences described in..." for clarity. **delete 'offer'**

AQ3. Please check if the edit made to the sentence "we select, first of all..." is fine. **OK, thanks**