EDITORS: Barry I. Schneider, bis@nist.gov
Gabriel A. Wainer, gwainer@sce.carleton.ca

DEPARTMENT: COMPUTER SIMULATIONS

Formal Modeling and Simulation for SARS-CoV-2 Containment Scenarios in Catalonia

P. Fonseca I Casas, J. Garcia I Subirana, and V. García I Carrasco, *Universitat Politècnica de Catalunya, 08034, Barcelona, Spain*

X. Pi I Palomés, *Universitat Oberta de Catalunya, 08018, Barcelona, Spain*

Gabriel Wainer, *Carleton University, Ottawa, ON, K1S 5B6, Canada*

We define interrelated models to simulate the spread of SARS-CoV-2 in Catalonia, which can be used to effectively build simulation applications and analyze the effects of nonpharmaceutical interventions. Due to the constant evolution of this pandemic, and the need to take a multidisciplinary approach, we use a formal specification to represent the model and to validate the model assumptions. We discuss the definition of the model using formal languages, and the Specification and Description Language to improve communication between stakeholders. We show formalization details, discuss implications in the validation process, and present how results obtained from the model of the pandemic in Catalonia can be used for decision-making.

Building computer simulations to study complex systems hinges on collaboration in multidisciplinary teams including experts with diverse backgrounds. The models need to consider the assumptions presented by the various specialists, which should be represented in the model to ensure that it reflects the system behavior accurately. During the pandemic of the disease caused by the SARS-CoV-2 virus, numerous simulation models were built and used to forecast the spread of the disease in different geographic regions. These models were built using a variety of methods, including differential equations, multiagent systems, Cellular Automata, and other alternatives. A common approach consists of using compartmental models, which have been available as early as 1930.[1] They divide the population under study into three main compartments: those including the population *Susceptible* to infection, those *Infected* and spreading the disease, and those *Removed* from the population under study (either because they are deceased or recovered). Numerous improvements to these *SIR* models have been considered, including the addition of new compartments leading to more realistic models. For instance, one can divide the Removed into two: *Deceased* and *Recovered*. One can also add an *Exposed* compartment, representing those that have been exposed to the disease (but do not spread the disease yet).

The changing situation of the COVID-19 pandemic made it important to be able to adapt the model as soon as new information about the disease became available. This changing situation also made it important to have models that could consider the influence of various nonpharmaceutical interventions (NPIs). And we needed to ensure that the simulation results were correlated with real-world data. Using a common language that all the experts involved in the process can understand, and ensuring the correctness of the assumptions made by the different experts is essential for the success of the effort. The team of experts can introduce errors when the model is built because of a misunderstanding of the system behavior, or because of errors introduced when implementing the computer programs that will execute the model. We improved such collaboration using formal models that provide an unambiguous common language for modeling, and

1521-9615 © 2022 IEEE
Digital Object Identifier 10.1109/MCSE.2022.3186227

a concrete mechanism to generate correct simulations. Conducting proper validation is difficult because of the numerous changes to the information about the disease. We use a Model Comparison Validation and Verification technique, which uses different models to improve error detection, both on definition and implementation: we define three models that represent the same system with different techniques to see if the results are similar, allowing the detection of errors. We combine three models that focus on different aspects of the model and can be used for cross validation between these models, increasing credibility of the forecast and simplifying error detection.

SD MODEL

The first model, seen in Figure 1(a) is a SEIRD[4] prototype defined using System Dynamics (SD) and focused on the analysis of the feasibility of the key assumptions used by the different experts. It is used to ensure that all main aspects are taken into consideration, for example, the composition of the population, the nature of the compartments to be used on the model, and the initial parametrizations for the SARS-CoV-2 spread. The evolution of an infected individual is Susceptible → Exposed → Infectious → Recovered or Dead. Each compartment represents a subset of the population in that specific state and the model of the pandemic evolves by transferring individuals from one compartment to another. More advanced versions of the SD model include other scenarios, like confinements (built by adding a path from Susceptible to Confined and allowing an analysis of the temporal evolution of the pandemic for each wave). This mode does not use real data, but it is used to analyze the preliminary model assumptions.

PYTHON MODEL

The second model, built using Python, uses the SD model structure and optimizes and fits the observed data using a simulated annealing algorithm2.[a] The objective is to obtain the parameters needed for the third model. The optimization model uses the observed cases as an input and tries to find a curve that fits with these observations (the mean of seven days). This model has the objective of studying the evolution of cases in Catalonia to define a basic transmission rate to be used in the SDL model. We estimate several parameters for the spreading of the virus, including the effective reproductive number R_t, which is used to measure the likely transmission of the disease. R_t represents the average number of secondary infections produced by a single infection (if $R_t>1$, the number of cases will increase, and if $R_t<1$, the number of cases will decrease). NPIs (e.g., reducing the mobility of the population, or mandating the use of a mask) affect the value of R_t. Our model estimates the Transmission Rate β (which is equivalent to R_t, calculated as $\beta = R_t\,\gamma$ with recovery rate γ). β is calculated using observed data that is inputted to our Python model, optimizing and fitting the observed data. We can find the trend of the observed cases, and when the trend changes (for instance, because of new interventions like vaccines), we can compute a new β value.

At this point, with the main assumptions made and the key parameters calculated, we build a third model that includes all the needed assumptions for decision making and that is able to provide a forecast based on historical data and expert knowledge.

SDL MODEL

The last model is built using Specification and Description Language (SDL),[2] which provides a formal and unambiguous mechanism to describe real-world systems.[3] The SDL model allows us to add detailed assumptions and behaviors when NPIs are applied or new knowledge about the virus spread is available. Our team developed a variety of versions of the SDL, each of which includes the assumptions or the interventions applied at the time of the analysis. To validate this third model, we compared the daily new cases forecast with a dataset that contains the daily new cases in Catalonia.[b]

SDL is a graphical object-oriented language with unambiguous formal semantics, standardized by the International Telecommunication Union. SDL uses four hierarchical building blocks: 1) SYSTEM, 2) BLOCK, 3) PROCESS, and 4) PROCEDURES. The SYSTEM and BLOCK diagrams represent the model's structure, using a hierarchical decomposition. PROCESS and PROCEDURES define the model's behavior. BLOCK and PROCESS are AGENTS that establish the communication, sending SIGNALS through CHANNELS. SIGNALS function as a trigger, generating the execution of a set of actions in a PROCESS. To represent time, all SIGNALS own a delay parameter and are sorted by delay and priority in the input queue of every input channel. SYSTEM is the topmost diagram in any SDL model, representing the main elements of the structure of the model.

Figure 1(b) shows the model used to forecast the second pandemic wave in Catalonia. The diagram

[a] Find the Python model in the CodeOcean service at the following URL: https://doi.org/10.24433/CO.9635632.v1

[b] http://governobert.gencat.cat/en/dades_obertes/

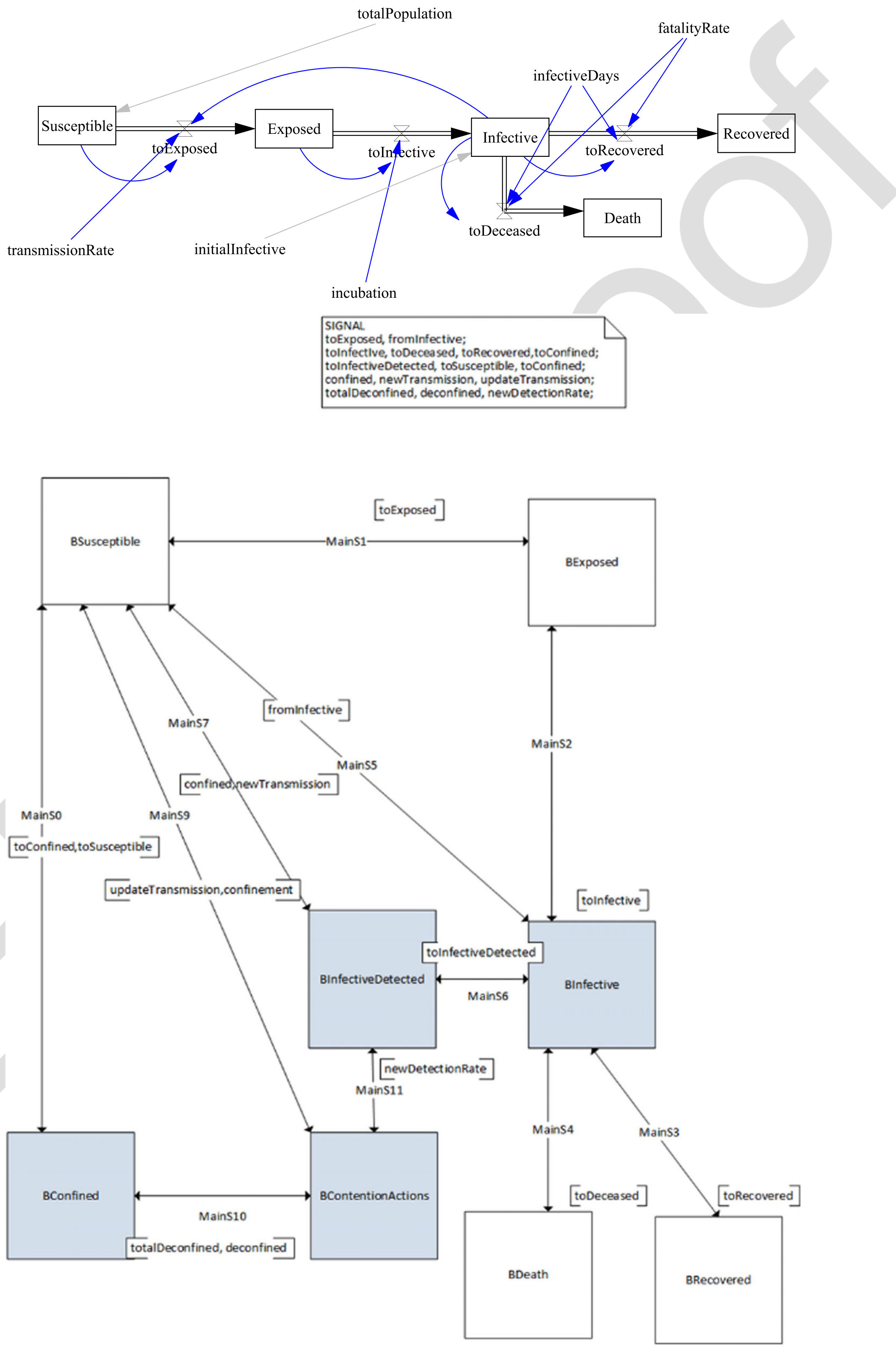

FIGURE 1. Model diagrams: on top, SD SEIRD model; bottom: SDL model SYSTEM diagram.

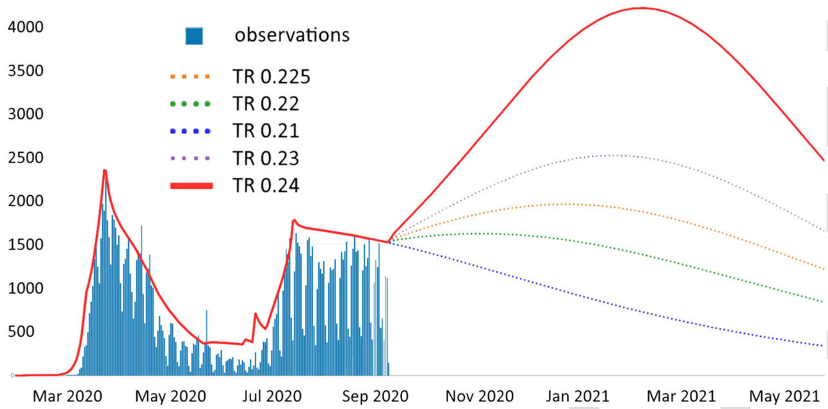

FIGURE 2. Forecast for new infections: red line, SDL model. Other scenarios with different hypotheses.

represents the SDL model SYSTEM diagram equivalent to the SD model defined earlier. On both diagrams, the squares represent the different compartments of the SEIRD model (BLOCKS in the SDL model, and by Levels for the SD model). We compare the outboth models' outputs for validation: the SD model provides an approximation to the pandemic's evolution, while the SDL model provides more detail, allowing us to understand the effects of the NPIs.

If needed, more BLOCKS can be added in the SDL model to include new behavior (for instance, vaccinated individuals, deceased, and other categories). These new BLOCKS of the SDL model are 1) *BConfinement*, representing the population in confinement that cannot become exposed; 2) *BInfectiveDetected*, the proportion of infected individuals detected, assuming that detection improves over time; 3) *BContainmentActions* controls the NPI actions taken by the authorities to prevent the spread of SARS-CoV-2.

The SDL model also includes a cellular automaton (CA)[5] in which each cell represents the spread of SARS-CoV-2 over each of Catalonia's Health Regions. This allows validating the model's assumptions by comparing the results with the data for each health region.

MODEL IMPLEMENTATION AND CALIBRATION

To implement the SDL models, we use SDLPS,[c] a computer software that can execute SDL and DEVS models.

To obtain accurate results, the models should fine-tune the simulation parameters, including the β parameter discussed earlier, to better predict the spread of the virus. To forecast the value of β during the second wave at the start of the school year in Fall 2020, we used common patterns found in other countries that started the school year earlier, such as South Korea (August 25, 2020) and Israel (September 1, 2020). Using the real-world data from those countries, we applied the Python optimization model to identify those data points where the β value changed. These were correlated with each of the NPIs put in place.

The simulation with the model of the second wave on Catalonia began on January 29, 2020. With the rise in cases, on the March 15, 2020, a lockdown was applied (except for essential workers). Air space was closed on the March 23. When the situation improved on the April 13, the workers of the industrial sector returned to work. On the April 20, the government provided basic masks to its citizens. In the following months, the restrictions were gradually lifted, until July 2020 At this point, we calculated a new β to represent a change in the spread of the virus using the analysis done on the data from Israel and used the cumulative incidence in Catalonia and Israel. We also considered that the numbers observed during the summer would increase due to a higher level of activity and contact, and that they would not be smaller than the numbers obtained during the lockdown. Other parameters were the detection rate and the asymptomatic population.[6] Figure 2 shows the forecast for new cases predicted using different scenarios.

The forecasting model uses past data to adjust the parameters to calculate the evolution curve of new cases. In the figure, the most likely situation is shown in red, along with other scenarios that represent less likely situations. This analysis of different options allows managers to understand the future evolution of the pandemic and make informed decisions. The red line in the figure shows the forecast most accepted by the domain experts considering all model assumptions and hypotheses. The simulation generates alternatives to help decision-makers when considering different NPIs (as well as errors in the experimental hypotheses). If during the evolution of the

[c]https://sdlps.com/

pandemic new available data modifies previous empirical results, the pandemic will evolve differently. The different trend lines shown on the figure can be useful to predict how to act in those cases. This can be used by domain experts in decision-making.

We built a dashboard[d] including Key Performance Indicators (KPIs) and the model forecast, which is continuously validated against real data (using a Continuous Solution validation approach). When the model gives invalid results (i.e., we detect a divergence between the model's forecast and real data time series), we reevaluate the hypotheses and recalibrate the model. This provides valuable information to understand the latest changes in the pandemic's evolution. This happens often since new NPIs are applied throughout the pandemic.

CONCLUSION

The definition of models and rapid prototyping is especially important to study the spread of diseases like COVID-19, caused by SARS-CoV-2. Simulation models allow building *what-if* scenarios to compare future possible trends of the pandemic evolution based on the past epidemics data. We showed how to build advanced models to improve the process, by building more than one formal model of the system under study (in our examples above, three related models) that allow continuous verification and validation of the model when new data from the system under study become available. The first model (using SD) allows us to understand the validity of the initial assumptions and the nature of the spread of the disease; a second model (in Python) allows studying the temporal behavior of the phenomenon and to estimate the parameters of the model. The third model (built in SDL) is used to understand the influence of NPIs on the population and perform the forecast. The combination of the three models provides a solid mechanism to validate the quality of the models and help make decisions and study the influence of NPIs. The use of formal modeling methods to represent the system improves communication with experts and analysis of the hypotheses while providing a mechanism for monitoring and continuous validation, improving decision making.

ACKNOWLEDGMENTS

This work was supported by CCD under Grant 2020-L015. The authors would like to thank inLab FIB for assistance with the project's administration.

[d]http://pand.sdlps.com

REFERENCES

1. R. Anderson, "The Kermack–McKendrick epidemic threshold theorem," *Bull. Math. Biol.*, vol. 53, pp. 3–32, 1991, doi: 10.1016/S0092-8240(05)80039-4.
2. E. Sherratt, I. Ober, E. Gaudin, P. Fonseca I Casas, and F. Kristoffersen, "SDL—The IoT language," *Int. SDL Forum.*, vol. 9369, pp. 27–41, 2015.
3. J. Leiva, P. Fonseca I Casas, and J. Ocaña, "Modeling anesthesia and pavilion surgical units in a Chilean hospital with specification and description language," *Simulation*, vol. 89, pp. 1020–1035, 2013, doi: 10.1177/0037549713495742.
4. E. Vynnycky and R. White, *An Introduction to Infectious Disease Modelling*. Oxford, U.K.: Oxford Univ. Press, 2010.
5. P. Fonseca I Casas, J. Garcia I Subirana, V. Garcia I Carrasco, J. L. Silva de Barcellos, J. Roma, and X. Pi, "SDL cellular automaton COVID-19 conceptualization," in *Proc. 12th Syst. Anal. Model.*, 2020, pp. 144–153.
6. H. Nishiura *et al.*, "Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19)," *Int. J. Infect. Dis.*, vol. 94, pp. 154–155, 2020, doi: 10.1016/j.ijid.2020.03.020.

P. FONSECA I CASAS is a professor at Polytechnic University of Catalonia (UPC), in the field of simulation, data processing, operations research, and statistics. Since 2017, he is a member of the committee 4.0 of the engineering industry in Catalonia. Contact him at pau@fib.upc.edu.

J. GARCIA I SUBIRANA is a professor at Polytechnic University of Catalonia (UPC) and Universitat de Barcelona, in the Statistics and Operations Research Department, with nearly 10 years of experience in projects related to modeling and simulation projects. Contact him at joan.garcia-subirana@upc.edu

V. GARCÍA I CARRASCO received the master's degree in computer and applied physics. He participates in simulation and modeling projects in Polyhedra Tech, a startup of Inlab FIB, and a TECNIO center of the Polytechnic University of Catalonia. Contact him at victor.garcia@polyhedra.tech.

X. PI I PALOMÉS teaches in the Master on Industry 4.0 at UPC, as well as Open University of Catalonia. He received the M.Sc. degree in industrial engineering from Polytechnic University of Catalonia (UPC). Contact him at xpi@enginyers.net.

GABRIEL WAINER is a professor with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada. He is a fellow of SCS, the Society for Modeling and Simulation International. Contact him at gwainer@sce.carleton.ca.