Future Generation Computer Systems 149 (2023) 547–561

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Evaluating the spread of Omicron COVID-19 variant in Spain☆

Miguel Guzman Merino[a], Maria-Cristina Marinescu[b], Alberto Cascajo[a,*], Jesus Carretero[a], David E. Singh[a]

[a] Universidad Carlos III de Madrid, Madrid, Spain
[b] Barcelona Supercomputing Center, Barcelona, Spain

ARTICLE INFO

Article history:
Received 3 July 2022
Received in revised form 22 May 2023
Accepted 20 July 2023
Available online 5 August 2023

Keywords:
Epidemiological simulation
COVID-19
Parallel simulator
Transportation

ABSTRACT

This work analyzes the propagation the highly transmissible COVID-19 variant Omicron across Spain via simulation by using EpiGraph. EpiGraph is an agent-based parallel simulator that reproduces the COVID-19 propagation over wide areas. In this work we consider a population of 19,574,086 individuals of the 63 most populated cities of Spain, for the time interval between May 15th 2021 and March 6th 2022. The main variants existing at the start of the simulation were the Alpha and Delta, with prevalence of 4% and 96%. Then, during the second half of November 2021, the Omicron variant appears in Spain. Due to the higher transmission of this new variant — about 2 times larger than Delta, it quickly spreads through all the cities and becomes the dominant strain in the country. In this work we analyze the propagation of this variant under different mobility restrictions and patient zero scenarios. We first define a baseline scenario which reproduces the existing conditions of the COVID-19 propagation in Spain for our period of study. We then consider alternative scenarios for different starting locations of the propagation. Finally, for each one of these scenarios, we evaluate different transportation intensities — i.e. movement of individuals between the cities. The main conclusion is that, independently of the initial location of the Omicron variant and the existing transportation conditions, the Omicron variant spreads through all the country in a short time interval. The work presented in this paper also implements and evaluates a power monitoring and optimization system aimed at reducing the energy consumption of such massive simulations as the ones performed in EpiGraph.

1. Introduction

During the first COVID-19 waves, transmission has been shown to diminish when following a combination of vaccination and non-pharmaceutical interventions such as the use of the face masks and social distancing measures. The degree of effectiveness of these measures is unclear for the more recent COVID-19 variants, which are much more contagious than the original strain. For instance, the Delta variant is about 1.8 times more transmissible than the original COVID-19 and the Omicron variant is nearly 2 times more transmissible than the Delta variant. In this work we use our EpiGraph simulator to evaluate the effectiveness of different policies meant to restrict the movement of people between different cities in Spain for the time interval comprised between May 2021 and March 2022. At this time the Delta and Omicron variants were dominant in Spain.

EpiGraph is an agent-based parallel simulator that models the propagation of influenza and COVID-19, including their variants. It connects various models that realistically reproduce the environment where the infection occurs, e.g. a highly-detailed social model, a vaccination model, or a transportation model, which modulate the inter-individual spread while each individual transitions through the states of an extended SEIR epidemic model. EpiGraph also considers non-pharmaceutical interventions, including population testing, use of mask and different degrees of social-distancing measures.

Agent-based approaches have the potential to model each individual's characteristics and interaction patterns, which can result in much more realistic simulations compared to other

☆ This work is part of the agreement between the Community of Madrid and the Carlos III University of Madrid for the Funding of research projects on SARS-CoV-2 and COVID-19 disease, project name "Multi-source and multi-method prediction to support COVID-19 policy decision making", which was supported with REACT-EU funds from the European regional development fund "a way of making Europe". This work was also funded by the Spanish Ministry of Science and Innovation Project New Data Intensive Computing Methods for High-End and Edge Computing Platforms (DECIDE). Ref. PID 2019-107858GB-I00 and by the project grant BCV-2021-3-0007 of the Spanish Supercomputing Network (RES).

* Corresponding author.
E-mail addresses: mguzman@pa.uc3m.es (M.G. Merino), maria.marinescu@bsc.es (M.-C. Marinescu), acascajo@inf.uc3m.es (A. Cascajo), jcarrete@inf.uc3m.es (J. Carretero), dexposit@inf.uc3m.es (D.E. Singh).

https://doi.org/10.1016/j.future.2023.07.025
0167-739X/

approaches [1,2]. One of the distinguishing features of EpiGraph is that it relies on realistic data for both individuals and their interaction patterns, which we extract by scaling from existing social networks and contact matrices. To realistically capture the temporal nature of interactions and the specifics or each profession or social activity, interconnections are time-dependent. EpiGraph takes different heterogeneous data sources as input by mapping them to the different parameters of the agent model [3].

This work includes three contributions. On one hand, we aim reproduce the propagation of the Delta and Omicron COVID-190's variants in Spain in order to contribute to a better understanding on the modeling for the epidemic simulation to coronavirus pandemic. We perform large-scale simulations, considering 63 different cities in Spain and 19,574,086 individuals. On the other hand, we analyze the impact of the movement of people between different cities on the variant propagation, considering both the existing conditions at this time in Spain, as well as other hypothetical scenarios with more restrictive social distancing measures. To model more realistically the movement of individuals between cities in post-COVID-19 conditions, we improve the transportation model used in EpiGraph. As a result of our experiments, we conclude that independently of the initial location of the Omicron variant and the existing transportation conditions, the Omicron variant spreads though all the country in a short time interval.

To be able to report statistically significant results we need to average over many simulations. We repeat the country-wide simulation 600 times; given that each simulation takes about 3 h and runs over 110 cores, in total we use about 200 Khours of CPU time. Our third contribution therefore targets the performance of EpiGraph, with specific focus on how to reduce the energy consumption of the execution. To do this, we use LIMITLESS – our scalable monitoring tool – to collect performance metrics for the compute nodes where EpiGraph is running. Based on these, LIMITLESS generates a profile that can be used to tune the Dynamic Voltage and Frequency Scaling (DVFS) dynamically to reduce energy consumption. DVFS is an extensively used technique for CPU power management, which allows fixing specific frequency/voltage for each core. We leverage this functionality to reduce the energy depending on the current resource requirements of the application.

The structure of this paper is the following: Section 2 provides a description of EpiGraph. The transportation model, tuned in this work for COVID-19, is presented in Section 3. Section 4 describes the simulation environment and the main results. Section 5 describes the integration of a monitoring tool for profiling the application, as well as the methodology we used to provide an energy-aware optimization by means of tuning the DVFS. Sections 6 and 7 present the related work and main conclusions of this work.

2. Background

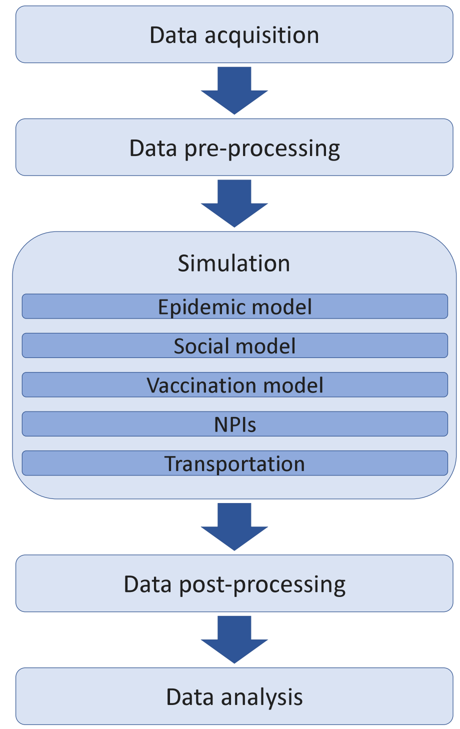


Fig. 1. Stages involved in EpiGraph simulator.

Fig. 1 shows the different stages involved in an EpiGraph simulation. The input data is first acquired from multiple sources ranging from research papers to public and private databases. These data are highly heterogeneous and have to be processed in a second pre-processing stage. The next stage corresponds to the simulation phase; the supplementary material provides details not included in this submission. When completed, it generates a collection of trace files with the state of each individual for each simulated time step in each urban area. This information is very rich in content and includes, in addition to the individual characteristics (health, age, occupation, etc.), the actions taken by or applied to the person (vaccination, use of NPIs, travel, etc.) at each time step. In order to provide a comprehensive analysis of the simulator results, a post-processing stage (fourth stage) is carried out. This stage generates specific trace files according the information that has to be analyzed or displayed. Finally, the fifth stage uses this information to generate statistical data that summarize the simulation output and graphically displays the results.

EpiGraph's structure is depicted in Algorithm 1. Each city simulates a population mix based on the Spanish census data [4], and defines social connections between its individuals. EpiGraph models each city as a labeled graph in which two individuals are connected by an edge if they are interconnected, with the label representing the connection type. This allows us to create time-dependent connections. For each simulation step (line 1) and every city in the simulated territory (line 2), the algorithm updates the health status of each infected individual as indicated by EpiGraph's Epidemic Model (line 5). The next step *ComputeSpreadGraph* (line 6) computes how the infectious pathogen spreads through the Social Model. We call non-pharmacological interventions applied at individual-level (*Individual_NPIs*, line 8) a non-pharmacological action taken by an individual to mitigate the propagation. An example of this kind of intervention is population sampling (testing) or using a surgical face mask at work but not during family time. In the case that sample testing is positive, the individual is quarantined, i.e. she is isolated from the rest of the population.

In line 10, dynamic transmissions are evaluated, as opposed to the static graph-based transmissions of line 6. Dynamic connections are generated for individuals belonging to certain collectives in which there are short-duration interactions with (many) different people. Examples of these collectives are health professionals and catering workers that are in contact with different patients or customers. These momentary connections change over the time, i.e. some of the individuals involved are different at every time step.

We call non-pharmacological interventions applied at collective-level (*Collective_NPIs*, line 11) an intervention – such as school closing or a total or partial lockdown – that is imposed (or lifted) by the health authorities at a certain time during the simulation and that involves a specific collective. Line 12 of Algorithm 1 simulates the exchange of individuals between cities by calling the Transportation Model. This model computes the number of travelers depending on the size of the origin and destination cities, as well as the geographical distance between them. Finally, in line 13 the Vaccination Model (depicted in [5]) captures both the COVID-19 vaccine availability and characteristics as well as the vaccination policies that determine which individuals are vaccinated at any time and with a certain vaccine type.

The following sections provide an outline of the different models used in the simulator.

Algorithm 1 EpiGraph transmission algorithm. Variable *simulation_time* represents the simulation duration, *simulated_territory* is the simulated area including several cities, each one of them, denoted as *city* with a social interaction model for the population. *Individual* contains characteristics and health status of each individual belonging to each city.

```
 1: for timestep = 1 → simulation_time do
 2:   for city ∈ simulated_territory do
 3:     for individual ∈ city do
 4:       if individual is infectious then
 5:         UpdateStatus(individual)
 6:         ComputeSpreadGraph(individual, city)
 7:       end if
 8:       Individual_NPIs(individual)
 9:     end for
10:     ComputeSpreadDynamic(city)
11:     Collective_NPIs(city)
12:     Transportation(city, simulated_territory)
13:     Vaccination(city)
14:   end for
15: end for
```

2.1. Social model

EpiGraph is an agent-based model that captures individuals with their attributes, affiliation to different groups of individuals, and interactions patterns. We consider four different main group types: students, workers, stay-at-home people, and elders. A group can represent a certain number of individuals that interact during work hours — for instance, groups represent students belonging to the same classroom, workers of the same company, or stay-at-home people and elderly people that perform group activities.

The way the individuals establish social contacts is time-dependent and reflects the temporal nature of the different types of interactions that each individual has throughout the day. For each one of the group types, we consider three different temporal distributions of the individual's activities, those related to weekdays, Saturdays, and holidays (including Sundays). These patterns are specific to the place being modeled to correctly capture typical work hours, school time, family time, and leisure. EpiGraph models interaction during all these time intervals, including specific interactions of several different professions. See [6,7] and Supplementary Material for additional details. Two different social-network graphs are used to generate the contact patterns of each individual. Additionally, contact matrices extracted from public surveys are used to provide statistical information of the average number of contacts between individuals of specific age ranges.

In order to increase the realism of population mixing, the work group has been broken down in different professions. We also distinguish different sub-collectives for the group of elderly people, separately for those living by themselves, attended in daily centers, or living in nursing homes. This work introduces ad-hoc connections, which are specific connection patterns for certain professions and collectives. These connections may be *static* or *dynamic*. Static connections are generated during the social model creation and do not change during the simulation. They are created when the connection graph is generated, before the simulator execution. Then, during the simulation these connections are evaluated by *ComputeSpreadGraph()* in Algorithm 1 (line 6), during certain time slots. Dynamic connections are generated during the simulation and change every time that they are evaluated. They reflect changing communication patterns of individuals who may be in contact with different individuals at different times.

ComputeSpreadDynamic() in Algorithm 1 (line 10) evaluates: (1) Ad-hoc school static connections, i.e. each educator is in contact with all the students of a certain class during work hours; (2) Ad-hoc elderly caregiver static connections, i.e. caregivers are in contact with a certain group of elderly people at a nursing home during work hours; (3) Ad-hoc catering dynamic connections, i.e. each worker belonging to this sector is in contact with clients during work hours; (4) Ad-hoc public security dynamic connections, i.e. the police and other security forces has contact with the general public; (5) Ad-hoc occasional meeting dynamic connections, that reflect meetings during social events.

Note that all the ad-hoc contacts are complementary to the existing graph-based contacts. This allows an individual to have two different types of interactions during work hours: stable connections with work colleagues (for instance, educators belonging to the same school, health professionals belonging to the same hospital or catering employees working at the same restaurant) but also ad-hoc connections with individuals in other group types (for instance, educators with students, health professional with patients and catering workers with customers).

2.2. Epidemic model

The epidemic model is a compartmental stochastic SEIR model extended with latent, asymptomatic, dead, hospitalized and vaccinated states. Rather than the more common analytic models based on differential equations, Epigraph probabilistically decides the duration of the different compartments and the transitions between them. In addition, the basic reproduction numbers R_0s are different for each compartment. It is an extended version of the figure presented in [8]. The different infection phases are described below:

- **Incubation stage**. At the beginning of this stage, individuals are infected but they have no symptoms and are not yet able to transmit the virus. This stage is represented as primary exposed E^P. From this stage the infection can enter one of two phases, based on a probability P^{EI}: a secondary exposed stage E^S where slight symptoms appear and the individual becomes infectious with a certain R_0^{ES}, or an asymptomatic stage.
- In the **asymptomatic stage** (compartment A), infected individuals do not notice symptoms but are able to transmit the disease with a certain R_0^A reproduction number. After a certain time, they pass to the recovered compartment in which the subject acquires viral immunity.
- In the first **symptomatic stage** – called primary infection state I^P – symptoms appear. Individuals will then transition to phase I^S, where symptoms persist. I^P, I^S and I^V have associated basic reproduction numbers of R_0^{IP}, R_0^{IS} and R_0^{IV}.
- A certain fraction of the individuals are hospitalized (**hospitalized stage**). The probability of entering this stage is given by the parameter $P^H(age)$, which increases with age.

From this state, an individual may transition to either the recovered or the dead stage. During hospitalization, we use R_0^H for modeling the transmission in hospitals.
- The individuals that reach the **dead stage** are removed from the simulation. The transition probability, $P^D(age)$, is also age-dependent and is applied over the portion of hospitalized individuals.
- The **treated stages** represent the infection stages for vaccinated individuals. A non-infected vaccinated individual is in the treated Susceptible state (S_T). If infected, he transitions to the treated Exposed primary (E_T^P). In case of a vaccination failure, the transition will include the states E_T^S, I_T^P, I_T^S and H_T, in a similar way to non-vaccinated individuals. Note that the probability of vaccination failure is p_T^I, which depends on the type of vaccine that has been used, the virus variant that has infected the individual, and, in some cases, other factors such as the age of the individual. If there is no vaccination failure the individual transitions to the Asymptomatic treated state A_T.

The time spent in a given state is generated following a normal distribution to simulate the time ranges specific to each stage of the infection and the fact that each individual may go through phases of different lengths. We also consider that a percentage of the sick individuals stay in bed, thus reducing the number of people that they interact with. We have used the same COVID-19 parameters (R0s values, transition probabilities, etc.) as the ones previously presented in [5,8].

It is worth mentioning that the different COVID-19 variants are modeled by assigning different values for the parameters of the epidemic and vaccination models (i.e vaccines have a different efficacy for each variant).

2.3. NPIs

The risk of infection, given by the specific R_0 value of the infected individual, also depends on two factors that reduce the transmission risk: the vaccination of the susceptible individual that is in contact with the infected one, and the use of non-pharmaceutical intervention (NPIs). This category includes different sources of heterogeneous intervention, including the following ones.

- Face mask use, that is related to the individual-based NPIs shown in Algorithm 1. EpiGraph models the use of both surgical and ffp2-grade face masks, with different efficacy [6].
- Social distancing policies are related to collective-based NPIs shown in Algorithm 1. It considers three distancing measures collected from the Data on country response measures to COVID-19 database [9]: closure of schools, closure of public spaces of any kind, and workplace closure. In this work we use the real social distance measures for Spain during the simulation period. We consider that during Christmas holidays all schools are closed, as well as a fraction of the workplaces; also, there are no restrictions in the use of public spaces or during leisure activities.
- Sampling strategies [7] are related to collective-based NPIs shown in Algorithm 1. These strategies are modeled by the number of daily tests that are performed, the minimum time between two consecutive tests carried out for the same individual, the quarantine time, and the percentage of quarantine breakers, i.e. the fraction of people who do not comply with social distancing during quarantine time. These data was provided by the Spanish Ministry of Health. Individuals that are COVID-19 positive in Spain are recommended to quarantine until at least three days after symptoms disappear and during at least a 10 days [10].

2.4. Vaccination model

In this study a generic COVID19 vaccine was modeled. In the compartment model used by EpiGraph, a vaccinated individual that is infected (E_T^P state) transitions to either the Asymptomatic Treated state (A_T) or – if the vaccine was not effective – to the Exposed Secondary Treated (E_T^S) state, then the infected treated states I_T^P and I_T^S, with equal probability of health risks (hospitalization (H_T) or death) as a non-vaccinated individual. The vaccination efficacy is modeled as the probability of transitioning to the A_T state. For instance, an individual vaccinated with a vaccine of efficacy 95% means that, if infected, he will have a 95% probability to transitioning to the A_T state.

EpiGraph currently models and simulates the Pfizer-BioNTech, Moderna, Astra-Zeneca and Janssen vaccines. The model considers multiple doses including a booster shot. We also consider the decay of the vaccine effectiveness depending on the vaccine type, individual characteristics (such as age), and the risk of COVID-19 reinfection among vaccinated and unvaccinated persons for the Omicron variant. [11] presents the EpiGraph's vaccination model; more details are also included in the supplemental material. The vaccination model includes both the vaccine effectiveness model that is subjected to waning and depends on the individual age and the SARS-CoV-2 variant, as well as the vaccination strategy that is simulated, which defines aspects such as prioritization among target groups and the time between the administration of the doses.

3. Transportation model

The transportation model reflects the movement of people between cities for work, study, or vacation, and it is based on the gravity model proposed by Viboud et al. [12]. Note that the movement of people within a city is already captured by the social model. The transportation model serves the purpose of moving individuals between different cities, allowing for disease transmission over large areas. The geographical information that EpiGraph takes into account includes latitude, longitude, and distance between urban regions, and was extracted from the Google Maps web service using the Google Distance Matrix API [13].

$$(d_{i,j} < 120 \text{ km}) \quad \Delta P_{i,j} = \frac{P_i^{0.30} Pj^{0.64}}{d_{i,j}^{3.05}} \tag{1}$$

$$(d_{i,j} \geq 120 \text{ km}) \quad \Delta P_{i,j} = \frac{P_i^{0.24} Pj^{0.14}}{d_{i,j}^{0.29}} \tag{2}$$

This model considers the exchange of individuals between cities, for each pair of cities i and j. This number ($\Delta Pi, j$) depends on the population size of the two locations (P_i and P_j) as well as the distance between them ($d_{i,j}$). Eq. (1) refers to travel distances of less than 120 km — which reflects the daily commute of students and workers to neighboring cities. Eq. (2) refers to the long-distance commute of workers that need to reside at a different location for several days in a row. The equations come from [12]. Additionally, we consider people from any group type that move at any distance for several days for vacation purposes. Once the volume of inter-city commuters is calculated, we randomly select individuals from specific group types within the populations and move them for a specific period of time to other locations. In our experiments, for the short distance commuters, 70% are workers and 15% are students. The remaining ones are elderly people and unemployed. For the long-distance commuters the percentages are 50% workers, 30% students, 15% retired individuals, and 5% unemployed people.

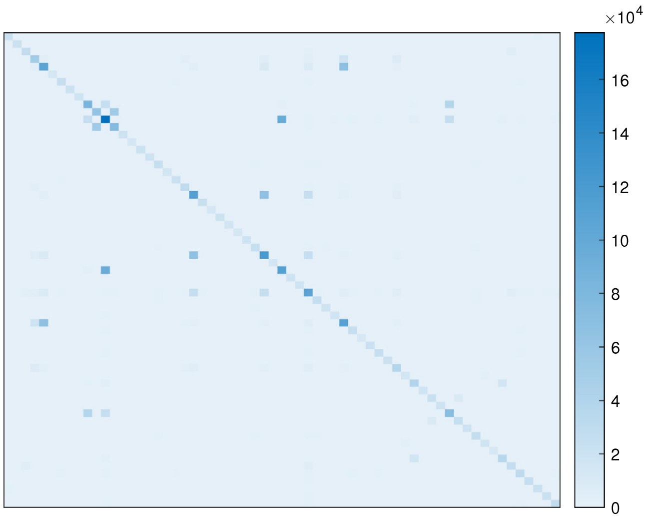


Fig. 2. Adjacency matrix A, of mobility values for the 63 cities considered in the simulation. X and Y axes corresponds to the different cities, thus each entry A[i][j] represents the number of inhabitants that transit from city *i* to city *j* per day. Note that in our model we approximate that A[i][j] = A[j][i]. The entry A[i][i] represents the overall number of individuals of city *i* that transit other cities.

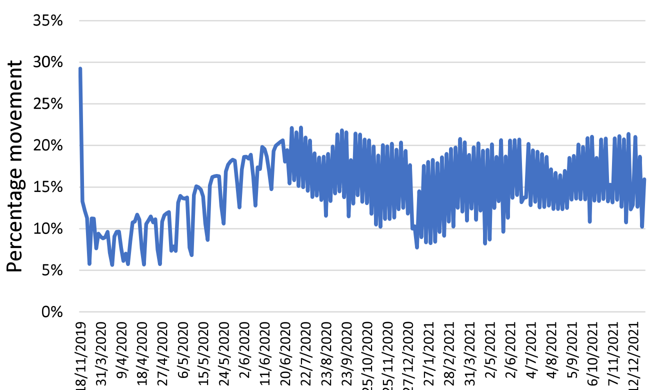


Fig. 3. Timeline of the average mobility percentages for Spain.

3.1. Extending the transportation model

The original transportation model was calibrated for a pre-pandemic situation. In this work we have introduced modification in the model in order to adapt it to the existing conditions in pandemic Spain. First of all, we scale down inter-city connections. Fig. 2 shows the adjacency matrix related to the baseline scenario. Cell A[i][j] represents the population of city i that travels to city j every day. Take Badalona and Barcelona, two nearby cities in 10th and 12th positions. Due to their large size and close distance, many individuals commute between them. A similar situation happens with Madrid (35th position) and the nearby satellite cities — e.g. Leganes or Getafe. The average percentage of mobility for the baseline scenario is 16.9%.

Since the transportation model takes into account both short-distance and long-distance connections (Eqs. (1) and (2)), populated cities separated by big distances (like Barcelona and Madrid) also exchange a significant amount of individuals. In this work, we have introduced a scale factor to these equations in order to reflect the reduction in transportation related to the COVID-19 pandemic (see Fig. 3). Besides the 16.9% of the base scenario, we have also considered three additional scenarios, which correspond to average mobility values of 2.5%, 11.4% and 20.9%, and we have evaluated the propagation of the Omicron variant in Spain in the time interval comprised between May 2021 and February 2022, for each of these four scenarios. These percentage values are chosen to be within the interval inferred from the mobility analysis we perform, described in the next subsection.

3.2. Mobility analysis

The mobility study [14] has been carried out in Spain before, during and after the first lock-down period in Spain, in Spring 2020. This study provides daily information on movements between the 3214 areas that were designed for the project and is based on information provided by telephone operators. For each mobile phone, the residence area is found as the one where the mobile phone is located for the longest time between 00:00 and 06:00 am during the last few months. This methodology approximates the number of terminals (i.e. mobile phones) that leave the residence area during the day. Fig. 3 shows the percentage of the population that leaves the residence area every day. We can observe that, due to the COVID-19 pandemic, there is a sharp decrease of the mobility at beginning of 2020. These values then increase but never reach their pre-pandemic values. Note that these values reflect the average values for Spain, but there are differences between the regions. For instance, on the 26th of November 2019 – before the lockdown –, the average national value was 26.5%, while the maximum value was 36.7% for Madrid. In contrast, on the 20th of June of 2020 – after the lockdown – the average value was 16.2%, with a value for Madrid of 18.8%.

4. Scenario simulation

4.1. Simulator configuration

The SARS-CoV-2 infection data were extracted from research papers. They include the basic reproduction numbers (R0s) related to each disease stage, the state transition probabilities (for instance, the probability of an infected individual of being asymptomatic), the hospitalized and death probabilities, and the duration of each infection stage. Please refer to [6] for a detailed description of these parameters.

We use the Spanish weekly sub-national 14-day notification rate of new COVID-19 cases [15] to set the initial percentage of infected population in each urban area (this value is only used at the beginning of the simulation). We obtain the seroprevalence information related to each autonomous community from [16]. This information is only needed to set the initial conditions of the simulation. The vaccination model was obtained from research papers and the vaccination strategy was provided by the Spanish Ministry of Health.[1] This model takes into account the booster shots and the vaccination of children. In our scenarios we assume that 30% of the children, (starting at 5 years old), 84% of the adults, and 96% of the elderly people have been vaccinated. The simulation considers a number of daily tests of 0.25% over the simulated population, and a percentage of positive tests of around 9% (which corresponds to the real testing rate and detection efficacy).

The Omicron variant was modeled to be 2 times more contagious than the Delta variant, but with a 70% smaller risk of developing dangerous symptoms. We assume a significant reduction of 30% in vaccine efficacy for preventing the infection with Omicron, but maintain the protection for developing dangerous symptoms. Finally, we consider a 8% risk of reinfection with Omicron variant among the individuals that have been previously infected with COVID-19.

[1] Note that the vaccination prioritization strategy is similar for all European countries.

4.2. Evaluation

In this section we provide simulation results for a national scenario related to Spain. We simulate the third wave starting on May 15th 2021. This scenario simulates a population of 19,574,086 individuals related to the 63 most populated cities of Spain, using 109 processes. The simulations were executed on the Tirant supercomputer, which is made up of 336 nodes, each of them with two Intel Xeon processors Sandy Bridge E5-2670 and 32 GB RAM, interconnected by an Infiniband 40 Gbps network. EpiGraph was compiled with Intel MPI. The execution time of each scenario takes a few hours with a memory footprint of roughly 18 GB of RAM.

The simulated Delta wave starts on May 15th of 2021 with a 0.8% of the population infected. The COVID-19 variant distribution is 4% and 96% prevalence for the Alpha and Delta variants. Then, around the second half of November 2021, the Omicron variant is introduced with a initial prevalence of 1%. Due to the higher transmissibility of Omicron, it quickly spreads throughout the country and becomes the dominant strain in Spain by the end of December 2021.

Fig. 4 shows the results for the baseline scenario in Spain, broken down by province. Real incidence, reported by [15] is displayed in red while the simulated values are displayed in blue. Note that the simulator considers all the active cases, including the reported, unreported and asymptomatic ones. In order to compare the results with the real cases, real incidence cases have been scaled in order to consider all the infected individuals (reported, un-reported and asymptomatic cases). The scale factor is 4, that reflects that for the Omicron wave only the 25% of the infections were reported. We can observe that the real and simulated values are similar, although there are some differences for some of the provinces which are related to the complexity of the modeling process. Note that for the baseline scenario (average transportation percentage of 16.9%) there is no single source of Omicron, but rather it is evenly distributed across all the cities at the beginning of the Omicron outbreak.

Fig. 5 shows the different percentages of COVID-19 variants during the simulation period. We can observe that at the time of its introduction, the Omicron variant quickly spreads and replaces Delta. Fig. 6 shows the distribution of infections, re-infections and hospitalizations broken down per age groups. The children (population under 18) is the collective with more infections, and the adults (population between 18 and 65) is the collective with a slightly larger percentage of re-infections than the children collective.

As a comparison point, we have also evaluated a different scenario in which only Barcelona has initial cases of Omicron. Barcelona was chosen because it is a large city – —1,950,561 inhabitants – but due to its less central placement it is less connected that other cities such as Madrid. Fig. 7 shows the incidence for this new scenario in which the 62 remaining cities are only (initially) infected through inter-city movement. Two different average transportation percentages are considered, one low (2.5%) and the other high (20.9%). We can observe that, despite the fact that there is only a single source of Omicron, this new variant quickly spreads to all the Spanish territory. The spread for the smaller mobility is (evidently) slighter slower, shown in Fig. 7 by the blue simulated curves which are a little shifted to the right compared to the baseline scenario. This effect does not happen when higher mobility values are simulated — green simulated curves. We have carried out experiments choosing other patient-zero cities, e.g. Madrid, with similar results.

Fig. 8 shows the (simulated) aggregated values for different mobility percentages for the Barcelona scenario, i.e. Fig. 8 represents the aggregated values of Fig. 7 with all the values of the subcharts merged into a single figure. Despite the introduction of severe restrictions in inter-city mobility, the Omicron variant is able to propagate between the cities obtaining similar incidences in all the scenarios.

Just like in Figs. 7 and 8 (blue) shows the slight shift of the infection peak for the lowest mobility. We can also observe that the final temporal distribution of the infection (i.e. the shape of the curves) is the same above a certain threshold of mobility of around 10% – Figs. 8 (green), (orange) and (magenta) –. This means that a highly transmissible disease, like the Omicron COVID-19 variant, is able to propagate across regions at the same rate even under moderate mobility restrictions.

Figs. 9, 10 and 11 compare the incidence when Omicron propagation starts exclusively in Madrid, Barcelona, or A Coruña. A Coruña was chosen for being a non-central city – like Barcelona – but a lot less populated at 244,850 inhabitants. We consider the least restrictive transportation scenario with an average mobility of 20.9%. We can observe that when comparing a central city (Madrid) with peripheric cities (Barcelona and a Coruña), there is again a small shift in the propagation peak. Based on our experiments, only near-zero mobility values – which is equivalent to a lockdown scenario – are able to avoid the Omicron to spread across the country.

Figs. 12 shows the effect of introducing another Delta-variant wave instead of an Omicron-variant wave. In this case the incidence is much smaller and the infection peak happens after the Omicron peaks due to the reduced transmissibility of the Delta variant.

5. Energy-aware optimization

Since 2021, our team has been providing the Spanish Ministry of Health with results on COVID-19 incidence. Being able to report statistically significant numbers requires repeating the simulations many times. In our tests, we performed about 200 K hours of simulations over about 110 cores, which implies a non-negligible amount of resources — especially in terms of energy. This created a strong motivation to reduce the energy consumption by providing an energy-aware environment that is able to optimize the application.

This section focuses on the execution environment, which includes LIMITLESS – a monitoring system which collects EpiGraph performance metrics – and Dynamic Voltage Frequency Scaling (DVFS), a technique for CPU power management. Executing applications at lower frequency/voltage reduces their power consumption but this may negatively affect the performance of the application due to reduced CPU speed. Our objective is to save energy while maintaining performance, which implies that we cannot reduce the power in CPU-intensive phases. To do this, we design energy-aware policies and we use profiling provided by LIMITLESS to set the DVFS at runtime according to the application phase that is being executed.

DVFS has been extensively used to reduce energy consumption at low CPU usage levels. Applications like EpiGraph combine CPU, Communication and I/O phases, which implies that the optimal CPU frequency can be different in these phases. Additionally, reducing the frequency may not necessarily reduce energy due to a (possibly linear) increase in execution time. In order to deal with this potential issue, we need to continue monitoring and profiling the application during the experimentation, identifying the best configuration.

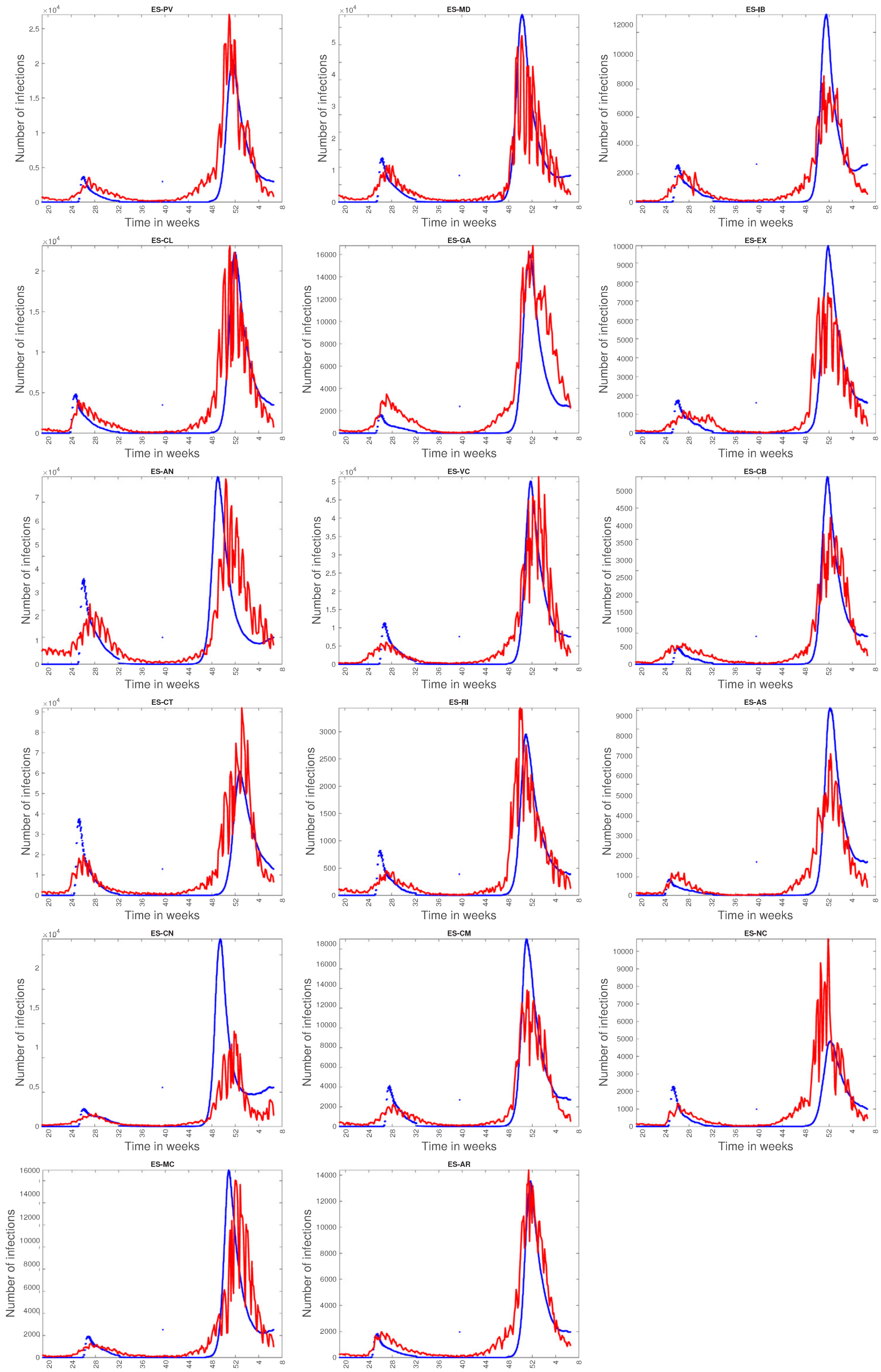


Fig. 4. Daily real (in red) and simulated (in blue) number of infections for the baseline scenario. The time interval we simulate is related to the Delta and Omicron waves that occurred between May 2021 (weeks 20–52) and February 2022 (weeks 1–8). Each one of the sub-charts are the results for a Spanish province. From top to bottom and left to right the provinces are País Vasco, Madrid, Islas Baleares, Castilla *y* León, Galicia, Extremadura, Andalucía, Valencia, Cantabria, Cataluña, La Rioja, Principado de Asturias, Islas Canarias, Castilla la Mancha, Comunidad Navarra, Murcia and Aragón.

5.1. Execution environment

LIghtweight MonIoring Tool for LargE Scale Systems (LIMITLESS) is a scalable monitoring system that can provide node-level performance information every 200 ms, although for this work we only collect it every second. LIMITLESS includes three main components (Fig. 13): a system monitor that collects the performance metrics from the cluster, an ElasticSearch [17] database for persistent storage, and LIMITLESS Analytics, that analyzes the executing applications and is responsible for the application modeling and profiling.

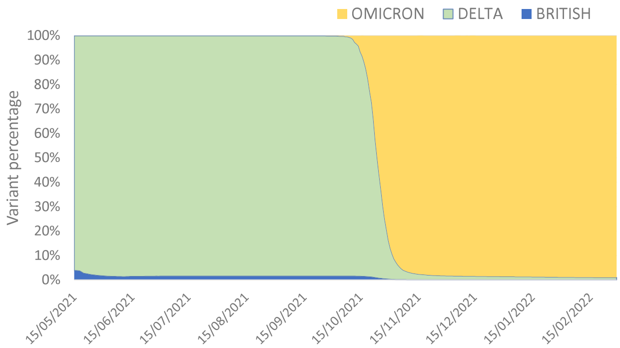


Fig. 5. COVID-19 variant percentage for the simulation period. In this time interval, the British, Delta and Omicron variants were the dominant strains.

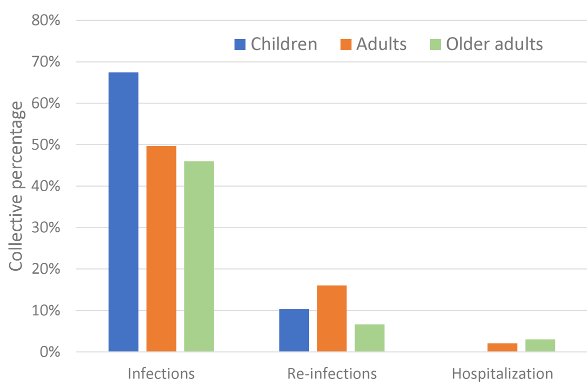


Fig. 6. Distribution of infections, re-infections and hospitalizations per collective. The y-axis represents the percentage of each collective that is infected, re-infected and hospitalized. Note that infections counts the total number of transmissions (including the re-infections) and re-infections only counts the individuals that were infected more than once.

During the EpiGraph simulations, LIMITLESS monitors every execution with their different DVFS values to generate application profiles. DVFS is configured at runtime depending on the information provided by the monitor. The time when DVFS updates take place is defined by the user and the environment executes them automatically; the overhead of DVFS reconfigurations is negligible.

The motivation behind generating application profiles is to identify if it is feasible to set different frequency/voltage values at runtime to reduce the energy consumption while maintaining the maximum expected performance. By executing an infection propagation scenario with different DFVS levels, we found that the maximum frequency level does not obtain the best global performance. It corroborates that the maximum CPU frequency/voltage is not always the best configuration, which rather depends on the applications. Besides, it demonstrates that knowing the behavior of applications it is possible to reduce the energy consumed while maintaining execution times (or even improving them) by experimenting with the DVFS.

5.2. Application profiling

Fig. 14 illustrates the methodology we propose to evaluate the application performance under different DVFS values. As a first stage, we characterize the application by collecting the performance metrics during its execution. The LIMITLESS system monitor component is in charge of providing these metrics. The second stage consists of storing the performance metrics associated with each test in the database, which allows the analytic component to generate the application profile. The third and last stage consists of dynamically updating the DVFS values during the application execution to consume less energy while maintaining its execution time.

The profiling process, which includes *application execution*, *performance metrics collection* and the *application profile generation*, is the first of this energy-aware optimization. In order to generate the application profiles, LIMITLESS provides performance data related to CPU, memory, I/O, communications, and current energy consumption. These performance metrics are processed by the analytic component in order to identify those application phases that are candidates to work with lower frequency and voltage. This methodology produces a set of performance-related metrics that are correlated between them to determine how to improve the application performance.

Fig. 15 shows the results of application profiling for the simulation of the infection wave during for Omicron COVID-19, in Madrid (Spain), during 2022. We can see one infection wave from day 700 until approximately day 820. We will call this time interval the *infection wave* while the rest conforms the *pre/post infection wave*. This figure is closely related to Fig. 16, which shows the time spent to compute each day of the simulation. Note that Fig. 15 starts on simulation day 548, but Fig. 16 takes into account every iteration of the application, since day one. So, we can see how infection waves impact the performance, increasing the iteration time at the time of active infections.

Fig. 17 shows the application profiling related to the I/O pattern during the simulation execution. As the main cause of I/O is checkpointing (EpiGraph performs checkpoints every 6 h), the expected behavior is to observe the same peaks during the entire execution. However, there is a reduction in the I/O peaks during the infection wave (when there are many active infections). This information leads us to consider that the performance behavior of the simulation changes during the infection wave, increasing the computation and iteration time and spacing out the I/O operations.

Figs. 18, 19, and 20 show the profiled CPU and memory use during the baseline simulation, as well as the network bandwidth, and the energy consumed. Sixteen processes are executed per node, consuming practically all the memory and maintaining a CPU consumption of nearly 80%. There is more network communication at the beginning of the execution due to the synchronization between the running processes. Once the simulation has been set up, the communications decrease to an average value of 400 Kbps (which is slightly lower during the infection wave). Finally, the energy is directly related to the selected frequency/voltage. Note that the default DVFS configuration sets the frequency/voltage at the highest level during the EpiGraph execution (2.2 GHz).

5.3. Evaluation

We run the evaluation on a configuration of six compute nodes with Intel(R) Xeon(R) Gold 6212U CPU @ 2.20 GHz, with 24 cores each and 315 GB of RAM. The connection between nodes is a 10 Gbps Ethernet and the I/O is based on Gluster parallel file system [18]. Given that main memory performance sees little or no degradation at reduced processor clock speeds, we allow memory use close to 100%. Guided by the profiling proposed by LIMITLESS, we propose a set of DVFS configurations for EpiGraph, which assign different frequencies for the infection wave and the pre/post infection waves. These configurations can be seen in Table 1 and Fig. 21.

The main result of these experiments is that the baseline, which uses the maximum frequency/voltage, does not provide

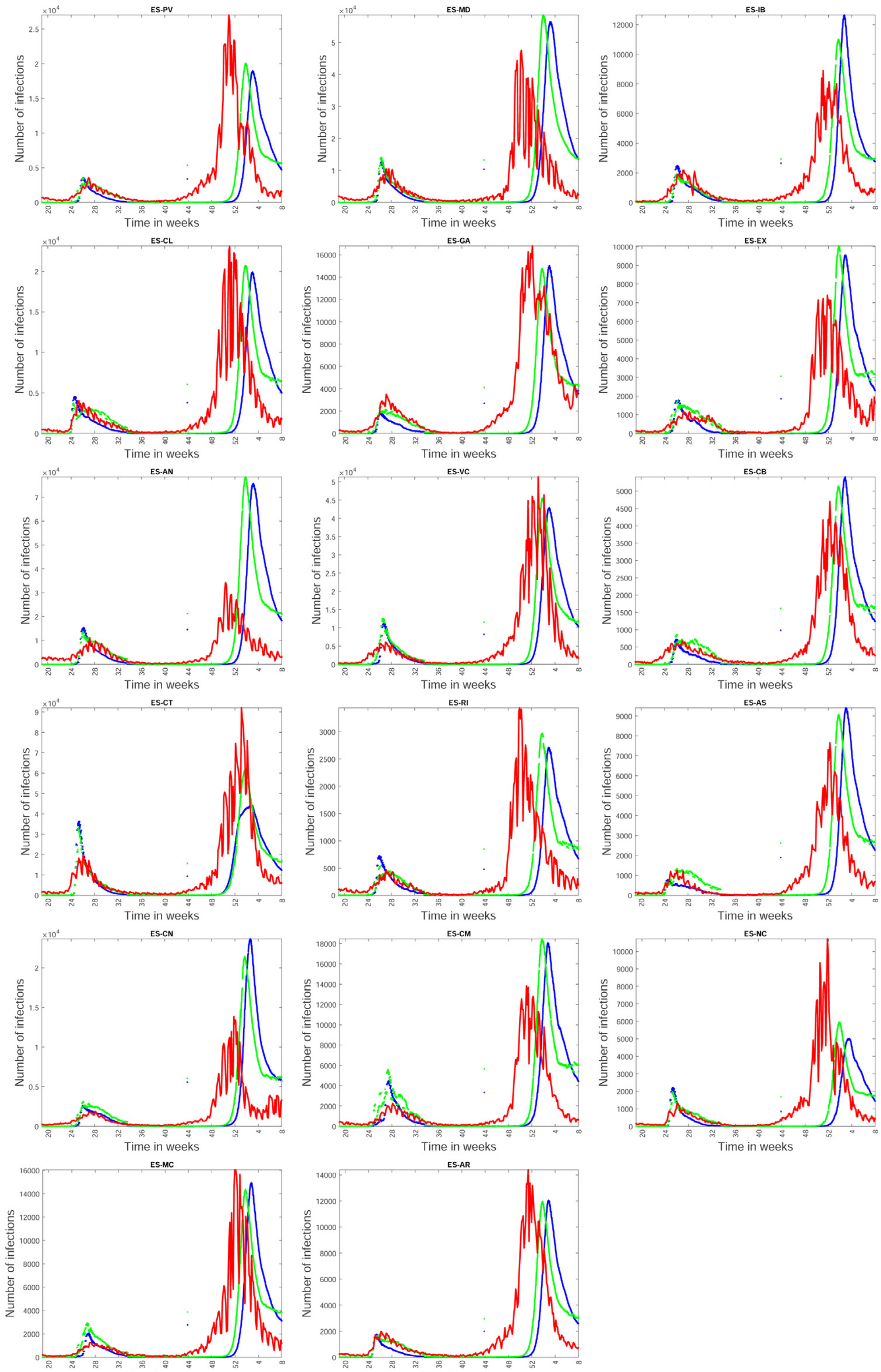


Fig. 7. New scenario where only Barcelona city has initial cases of Omicron variant: (blue) the average mobility of 2.5%, (green) average mobility of 20.9%. The time interval is related to the Delta and Omicron waves that occurred between May 2021 (weeks 20–52) and February 2022 (weeks 1–8). Each one of the sub-charts are the results for a Spanish province.

the best relation between performance and power consumption. Executing EpiGraph with a frequency of 2.2 GHz (the maximum frequency allowed by the architecture) requires 7540 s and consumes 874,805 W. Since the CPU speed is directly affected by the DVFS value and the memory is largely unaffected [19], the checkpointing is the likely culprit for the performance degradation at high frequency/voltage values.

Figs. 22, 23, and 24 show the performance metrics related to the test ID 20 in Table 1, which has a lower execution time and energy consumption than the baseline. The default frequency for the application is 1.9 GHz, while during the infection wave, the frequency is fixed at 2.2 GHz (maximum). This test needs 6905 s and 766,716 W to complete. If we compare the figures with the baseline, the, Fig. 22 has minimum CPU levels that are almost 20%

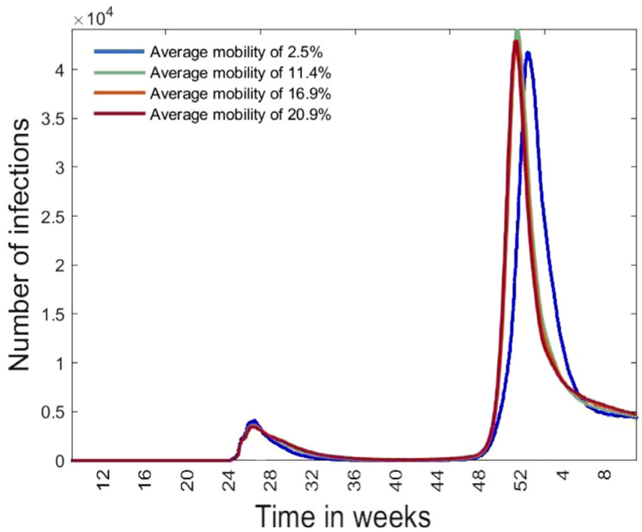


Fig. 8. Simulation aggregated results for a new scenario where only Barcelona city has initial cases of Omicron variant. Different mobility values, represented as the percentage of the city's population that transit to another city.

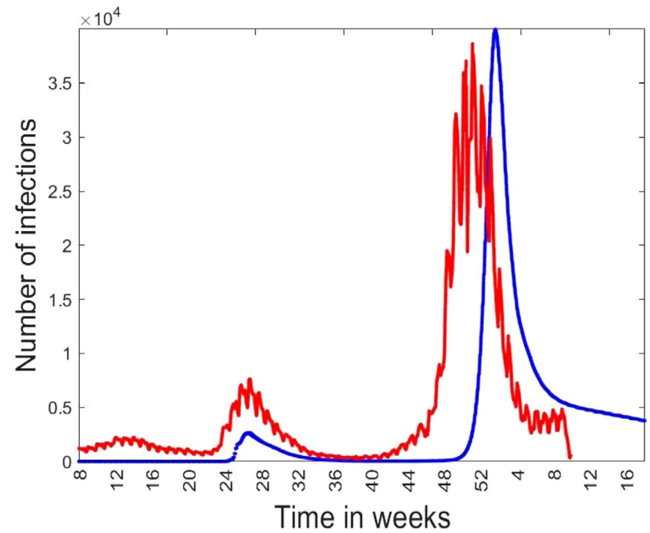


Fig. 10. Simulation aggregated results (in blue) for a new scenario where only Barcelona city has initial cases of Omicron variant. Official reported cases are displayed in red. Average mobility is 20.9%.

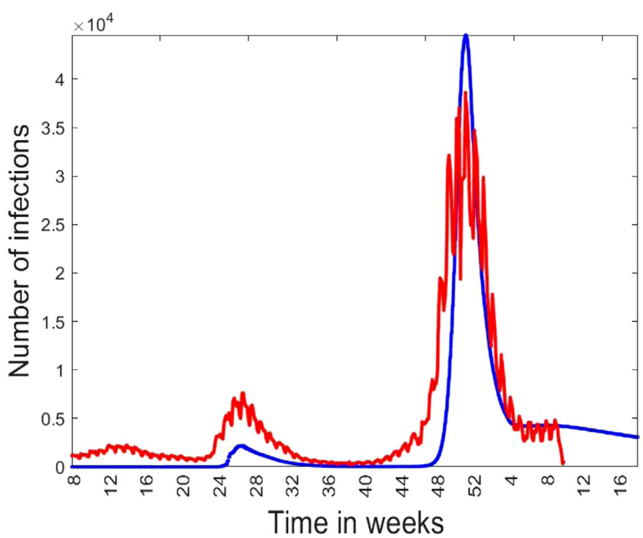


Fig. 9. Simulation aggregated results (in blue) for a new scenario where only Madrid city has initial cases of Omicron variant. Official reported cases are displayed in red. Average mobility is 20.9%.

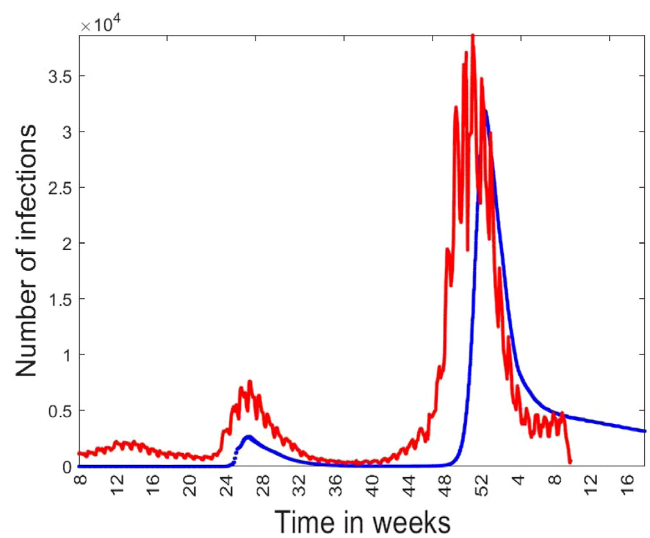


Fig. 11. Simulation aggregated results (in blue) for a new scenario where only A Coruña city has initial cases of Omicron variant. Official reported cases are displayed in red. Average mobility is 20.9%.

better. This contributes to an increase in the performance of the application. The same reasoning can be applied to Fig. 23, where the communication bandwidth reaches bigger values at the beginning of the application execution. Finally, Fig. 24 shows the energy consumed during the simulation. Different from Fig. 20 (baseline), the energy consumption is smaller before the infection wave. Compared to the baseline, reducing the frequency for the setup, initial, and final phases results in a reduction of the execution time of 8.4% and of the energy consumed by 12.3%.

6. Related work

There are many approaches to model the COVID-19 propagation. A starting approach is the SEIR model based on solving the differential equations like in [20]. More complex versions of the SEIR model include, for instance, a quarantine class and a class of isolated (hospitalized) members [21,22]. The main limitation of this approach is the lack of details in the simulation. An alternative way of modeling the infection spread are the models based on machine learning [23]. The work in [24], developed in the Imperial College of London introduces an extension of a semi-mechanistic Bayesian hierarchical model that infers the impact of interventions and estimates the number of infections over time. In [25], the authors use the discrete renewal equation as a latent process for the modeling of infections and propose a generative mechanism to connect infections to death data. They use this joint Bayesian hierarchical model to produce short-term predictions, and they apply their model to 11 different countries.

The European Centre for Disease Prevention and Control (ECDC) [26] has built a Monte-Carlo based model of COVID-19 that they use for forecasting. To model the behavior of the people and how well they are responding to the measures, they compare the predictions with Google data about mobile phone use and they use the daily confirmed COVID-19 cases and daily deaths to calibrate it. It is interesting to note that some models perform

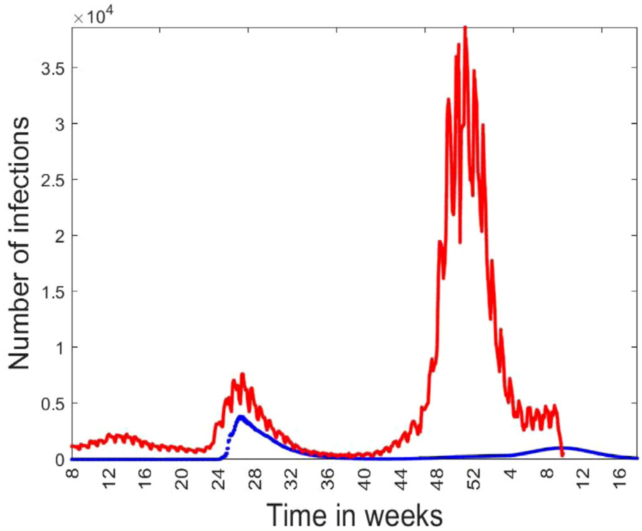


Fig. 12. Simulation aggregated results (in blue) for a new scenario where only Madrid city has initial cases of Delta variant (instead of Omicron). Official reported cases are displayed in red. Average mobility is 20.9%.

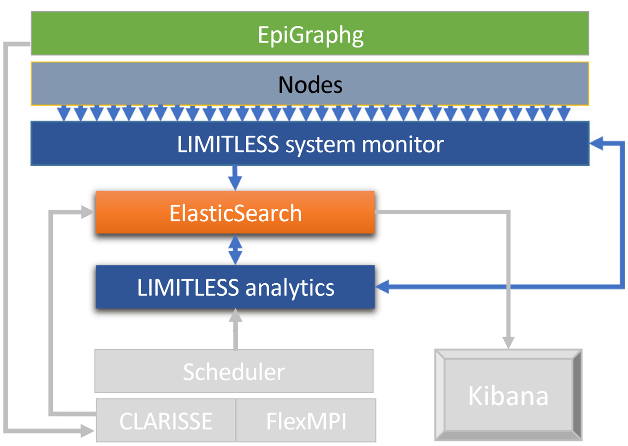


Fig. 13. LIMITLESS framework architecture. The monitor collects performance metrics from the nodes running EpiGraph, stores them into an ElasticSearch database, and processes them to create a profile of the application.

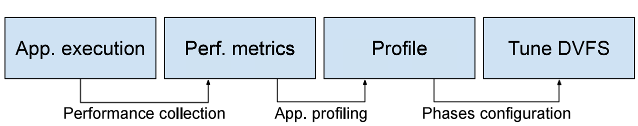


Fig. 14. Methodology to create the profiles based on the monitored data.

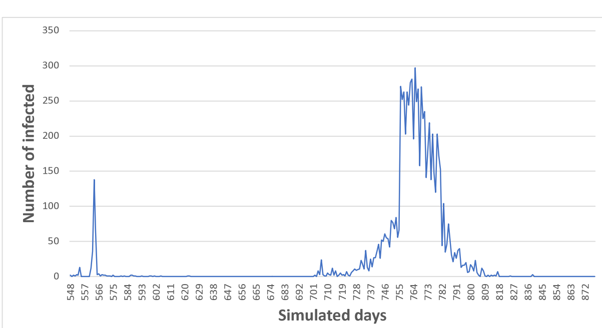


Fig. 15. Number of infected people per simulated day. This infection wave corresponds to the use case of Madrid.

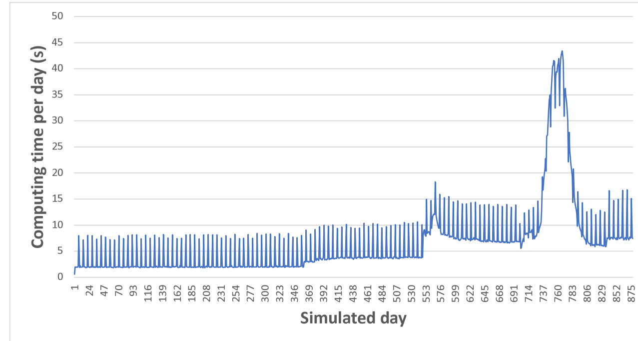


Fig. 16. Aggregated iteration time per simulated day. Every simulated day consists of 1440 iterations. As the iteration time depends on the number of the existing infected individuals but also on checkpointing, and communication between the processes, each simulated day has a different computation time.

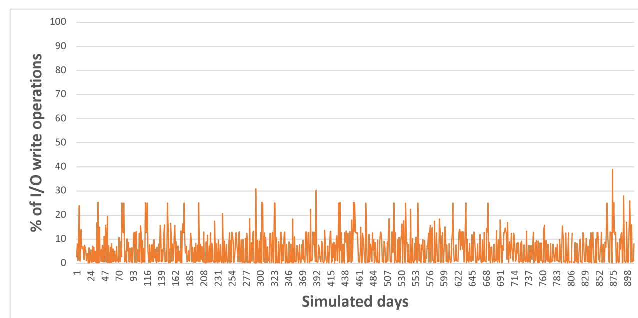


Fig. 17. I/O pattern during the use case of Madrid. Most of the I/O operations correspond to checkpointing operations.

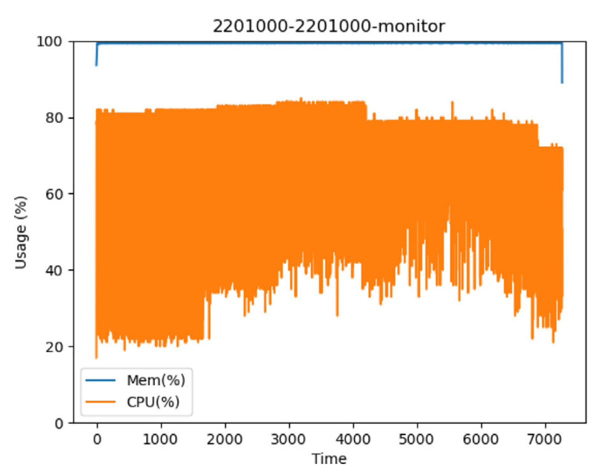


Fig. 18. EpiGraph baseline — CPU and memory usage using the DVFS fixed at a frequency at 2.2 GHz.

forecast, like COFFEE model from Los Alamos National Laboratory [27], and other are also capable of performing projections. A projection involves simulating alternative hypothetical scenarios. In the case of EpiGraph, this tool belongs to the models that perform projection.

Studies of the Delta variant before widespread booster vaccination are mixed on whether SARS-CoV-2 breakthrough infections in vaccinated individuals are potentially less infectious [28–30] or equally infectious [31,32] to primary infections. In more recent household contact studies during the Omicron variant wave [33–35], vaccination seems to lead to reduced SARS-CoV-2 infectiousness. In [36], the authors report on the infectiousness of

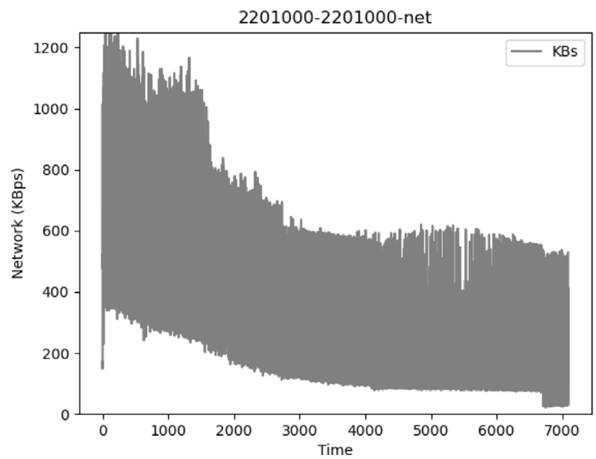


Fig. 19. EpiGraph baseline — Network usage using the DVFS fixed at a frequency of 2.2 GHz.

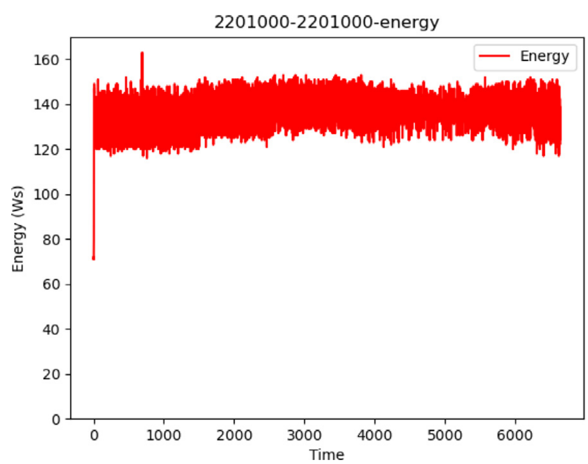


Fig. 20. EpiGraph baseline — Energy consumed using the DVFS fixed at a frequency of 2.2 GHz.

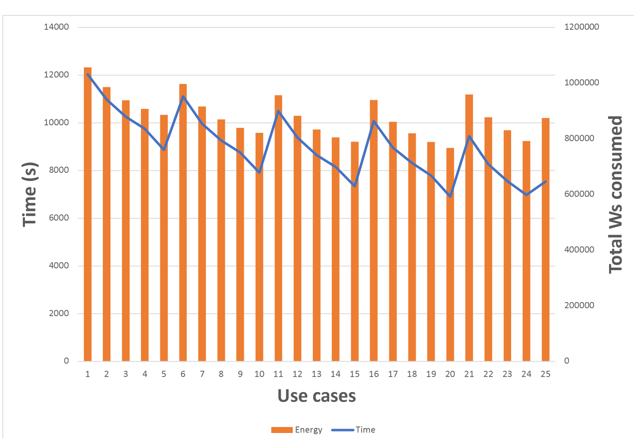


Fig. 21. DVFS combinations — For each frequency/voltage combination identified in Table 1, this figure shows the energy consumption as orange bars and the execution time as a blue line.

Table 1
Combination of frequencies/voltages for the two main identified application phases (pre/post infection wave and infection wave). For each test, the first value corresponds to the frequency established for the pre/post-infection wave and the second one fixes the frequency during the infection wave. These experiments have been executed three times in order to avoid occasional discrepancies.

ID	Use case	Exec. time (s)	Energy (Ws)
1	1 000 000–1 000 000	12 029	1 055 995
2	1 000 000–1 300 000	10 978	985 835
3	1 000 000–1 600 000	10 254	937 910
4	1 000 000–1 900 000	9750	907 560
5	1 000 000–2 201 000	8865	886 050
6	1 300 000–1 000 000	11 107	997 202
7	1 300 000–1 300 000	9963	915 690
8	1 300 000–1 600 000	9262	868 980
9	1 300 000–1 900 000	8758	839 613
10	1 300 000–2 201 000	7916	821 510
11	1 600 000–1 000 000	10 491	956 779
12	1 600 000–1 300 000	9383	882 455
13	1 600 000–1 600 000	8648	833 345
14	1 600 000–1 900 000	8153	805 137
15	1 600 000–2 201 000	7336	789 528
16	1 900 000–1 000 000	10 074	939 404
17	1 900 000–1 300 000	8955	860 864
18	1 900 000–1 600 000	8313	819 558
19	1 900 000–1 900 000	7795	788 504
20	**1 900 000–2 201 000**	**6905**	**766 716**
21	2 201 000–1 000 000	9440	959 051
22	2 201 000–1 300 000	8262	877 059
23	2 201 000–1 600 000	7554	830 169
24	2 201 000–1 900 000	6973	791 633
25	2 201 000–2 201 000	7540	874 805

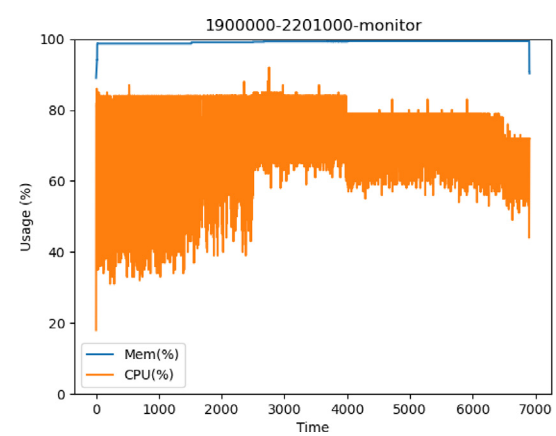


Fig. 22. EpiGraph use case 20 — CPU and memory usage using the DVFS with a frequency of 1.9 GHz, except between the simulated days 700 and 836, configured for 2.2 GHz.

SARS-CoV-2 infections occurring in vaccinated individuals and/or those with prior infection relative to unvaccinated and previously uninfected individuals who were incarcerated in a US state prison system during the first 5 months of the Omicron wave. They show that, irrespective of vaccination and/or prior natural infection, SARS-CoV-2 breakthrough infections and reinfections remained highly infectious and were responsible for 80% of transmission observed in the population under study, with high levels of both prior infection and vaccination. The implication is that, by themselves, vaccination and prevalent naturally acquired immunity will not eliminate risk of SARS-CoV-2 infection, although these results apply for the special case of high density locations without much internal mobility nor population exchange with other locations. The public health implication of these findings support the policy using booster doses of vaccination to lower transmission.

The related work related to energy optimization can be divided into two different points of view: system-level and application-level methods. Mair et al. [37] quantified the energy efficiency of the supercomputers using the Top500 and Green500 lists. They proposed a metric that weighs system scale and performance to evaluate power efficiency. One of the main conclusions,

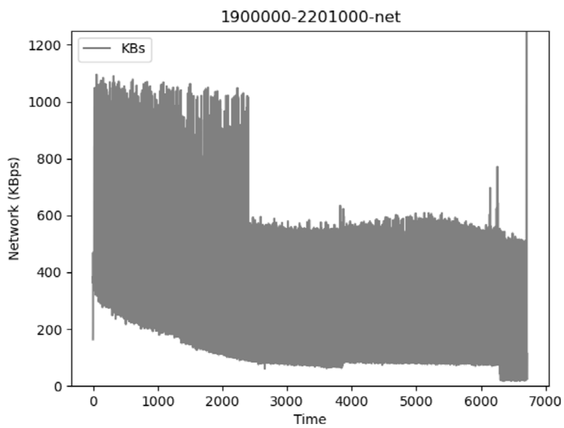


Fig. 23. EpiGraph use case 20 — Network usage using the DVFS with a frequency of 1.9 GHz, except between the simulated days 700 and 836, configured for 2.2 GHz.

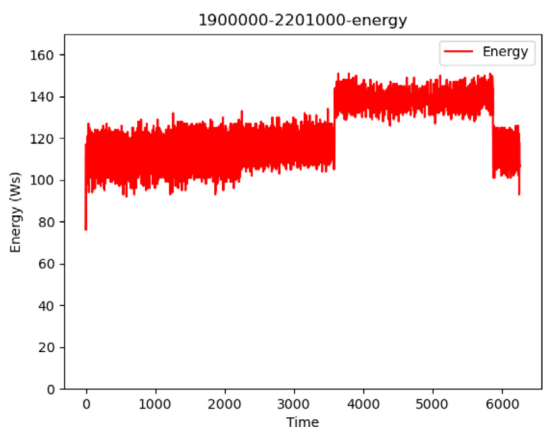


Fig. 24. EpiGraph use case 20 — Energy consumed using the DVFS with a frequency of 1.9 GHz, except between the simulated days 700 and 836, configured for 2.2 GHz.

which puts the focus on the DVFS and the energy-aware algorithms due to Exascale, is that the most efficient platforms are small-scale systems, confirming that large-scale systems rarely take into account energy consumption.

An energy-efficient algorithm using per-core DVFS with an adaptive runtime system is presented in [38]. The authors identify a set of three categories of applications that could benefit from their solution. However, they test their proposal using a set of micro-benchmarks and a framework that allows the developer to identify the different phases of their applications and thus apply the DVFS changes in a local manner. LIMITLESS creates an application profile to evaluate the performance instead of segmenting the code to apply local changes, which makes our alternative more generic. The authors also discuss the limitations of using DVFS in HPC environments (for example, DVFS and HyperThreading are incompatible because two threads can share the same physical core). Bratek et al. [39] study the benefits of using dynamic DVFS to improve the power performance of parallel applications on multi-core systems. The authors propose a methodology to adapt the frequency/voltage dynamically for each core in use depending on the work it performs. However, this optimization requires a more sophisticated algorithm to manage the different information about the processes and cores. It consumes more resources and requires continuous monitoring of the application and load of every core used by that application. Following this research line, Gupta et al. [40] introduce an algorithm to test different DVFS configurations based on a regression model which increases savings by more than 20% compared to executing without frequency tuning. However, the authors base their conclusions on simulation evaluations, while they do obtain better results than related works that are also based on simulation. It is important to say that linear regression-based models are often less accurate in predicting metrics than other machine learning models in case of high variability of the metrics — which is the case for HPC applications [41,42].

7. Conclusion

In this work we use the EpiGraph simulator to evaluate the propagation of the Omicron COVID-19's variant for Spain – for which we consider 63 cities and 19,574,086 individuals – for the time interval comprised between May of 2021 and March of 2022. EpiGraph integrates different models that reproduce the existing conditions in Spain, including the use of face mask, population testing, vaccination, and social distancing. In our study, we analyze multiple initial locations of this variant and different levels of movement of individuals between the cities. To reduce power consumption and execution time during the execution of the massive simulations we are running, we implement a monitoring and optimization system which is able to profile the applications and update the CPU frequency/voltage dynamically depending on the profiles. We demonstrate that using the default DVFS configuration does not always provide the best performance for the applications. In the case of EpiGraph, reducing the frequency when the simulation is not computing infection waves reduces the execution time by 8.4% and the energy consumed by 12.3% compared with the baseline (maximum CPU frequency allowed).

The main conclusion of this work is that, independently of the initial location of the Omicron variant, and the existing transportation conditions, the high transmissibility of Omicron variant – about 2 times larger than Delta, and roughly the 3.6 times more transmissible than the initial COVID-19 strain –, allows it to quickly spread throughout the country and become the dominant strain independently of the initial conditions of the simulation. The application optimization we have implemented makes running massive simulations for wide areas (e.g. European level) more sustainable in terms of time and resource utilization.

CRediT authorship contribution statement

Miguel Guzman Merino: Conceptualization, methodology, Software, Investigation, Formal analysis, Data curation, Writing – original draft, Visualization. **Maria-Cristina Marinescu:** Conceptualization, Investigation, Writing – original draft, Visualization, Review & editing. **Alberto Cascajo:** Conceptualization, Methodology, Software, Investigation, Data curation, Validation, Writing – original draft, Review & editing. **Jesus Carretero:** Conceptualization, Supervision, Funding acquisition. **David E. Singh:** Conceptualization, Methodology, Software, Investigation, Formal analysis, Data curation, Writing – original draft, Visualization, Validation, Review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.future.2023.07.025.

References

[1] Robert Hinch, William JM Probert, Anel Nurtay, Michelle Kendall, Chris Wymatt, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, et al., Openabm-COVID19-an agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing, medRxiv (2020).

[2] Erik Cuevas, An agent-based model to evaluate the COVID-19 transmission risks in facilities, Comput. Biol. Med. 121 (2020) 103827, http://dx.doi.org/10.1016/j.compbiomed.2020.103827, URL https://www.sciencedirect.com/science/article/pii/S001048252030192X.

[3] Miguel Guzmán Merino, Christian Duran Gonzalez, María Cristina Marinescu, Concepción Delgado Sanz, Diana Gómez Barroso, Jesús Carretero Pérez, David Expósito Singh, Data management in epigraph COVID-19 epidemic simulator, in: Workshop on 27th International European Conference on Parallel and Distributed Computing (Europar 2021)., 2021.

[4] National statistics institute (INE), 2021, http://www.ine.es.

[5] David E. Singh, Maria-Cristina Marinescu, Miguel Guzmán-Merino, Christian Durán, Concepción Delgado-Sanz, Diana Gomez-Barroso, Jesus Carretero, Evaluation of vaccination strategies for Spain, Preprint (2021).

[6] David E. Singh, Maria-Cristina Marinescu, Miguel Guzmán-Merino, Christian Durán, Concepción Delgado-Sanz, Diana Gomez-Barroso, Jesus Carretero, Simulation of COVID-19 propagation scenarios in the madrid metropolitan area, Front. Public Health 9 (2021) 172, http://dx.doi.org/10.3389/fpubh.2021.636023, URL https://www.frontiersin.org/article/10.3389/fpubh.2021.636023.

[7] Miguel Guzmán-Merino, Durán, Maria-Cristina Marinescu, Christian, Concepción Delgado-Sanz, Diana Gomez-Barroso, Jesus Carretero, David E. Singh, Assessing population-sampling strategies for reducing the COVID-19 incidence, Preprint (2021).

[8] David E. Singh, Maria-Cristina Marinescu, Miguel Guzmán-Merino, Christian Durán, Concepción Delgado-Sanz, Diana Gomez-Barroso, Jesus Carretero, Simulation of COVID-19 propagation scenarios in the madrid metropolitan area, Front. Public Health 9 (2021) 172, http://dx.doi.org/10.3389/fpubh.2021.636023, URL https://www.frontiersin.org/article/10.3389/fpubh.2021.636023.

[9] European Centre for Disease Prevention and Control(ECDC), Home-ECDE, 2021, https://www.ecdc.europa.eu/en.

[10] Ministerio de sanidad de España, Instituto de salud Carlos III (ISCIII), ESTRATEGIA DE DETECCIÒN PRECOZ, VIGILANCIA Y CONTROL DE COVID-19.

[11] David E Singh, Carmen Olmedo Luceron, Aurora Limia Sanchez, Miguel Guzman Merino, Christian Duran Gonzalez, Concepcion Delgado-Sanz, Diana Gomez-Barroso, Jesus Carretero, Maria-Cristina Marinescu, Evaluation of vaccination strategies for the metropolitan area of madrid via agent-based simulation, BMJ Open 12 (12) (2022) http://dx.doi.org/10.1136/bmjopen-2022-065937, arXiv:https://bmjopen.bmj.com/content/12/12/e065937.full.pdf, URL https://bmjopen.bmj.com/content/12/12/e065937.

[12] Cécile Viboud, Ottar N Bjø rnstad, David L Smith, Lone Simonsen, Mark A Miller, Bryan T Grenfell, Synchrony, waves, and spatial hierarchies in the spread of influenza, Science 312 (5772) (2006) 447–451.

[13] Google maps API, 2021, https://developers.google.com/maps/.

[14] Subdirección General de Estadísticas Sociodemográficas, Estudios de movilidad de la población a partir de la telefonía móvil 2020–2021, 2021, https://www.ine.es/experimental/movilidad/exp_em_proyecto.pdf.

[15] SiViES (Sistema de Vigilancia de España), Incidencias acumuladas e indicadores de transmisibilidad, 2022, https://cnecovid.isciii.es/covid19/.

[16] Ali Rostami, Mahdi Sepidarkish, Mariska Leeflang, Seyed Mohammad Riahi, Malihe Nourollahpour Shiadeh, Sahar Esfandyari, Ali H Mokdad, Peter J. Hotez, Robin B. Gasser, First "snap-shot" meta-analysis to estimate the prevalence of serum antibodies to SARS-CoV-2 in humans, Clin. Microbiol. Infect. 27 (3) (2021) 331–340.

[17] What is ElasticSearch? | Elastic, https://www.elastic.co/es/what-is/elasticsearch.

[18] The Gluster file system., http://www.gluster.org/.

[19] Wen-Yew Liang, Shih-Chang Chen, Yang-Lang Chang, Jyh-Perng Fang, Memory-aware dynamic voltage and frequency prediction for portable devices, in: Embedded and Real-Time Computing Systems and Applications, 2008. RTCSA '08. 14th IEEE International Conference on, 2008, pp. 229–236, http://dx.doi.org/10.1109/RTCSA.2008.19.

[20] José M. Carcione, Juan E. Santos, Claudio Bagaini, Jing Ba, A simulation of a COVID-19 epidemic based on a deterministic seir model, Front. Public Health (2020).

[21] F. Brauer, C. Castillo-Chavez, Mathematical Models in Population Biology and Epidemiology, Springer, 2012.

[22] Michael Lingzhi Li, Hamza Tazi Bouardi, Omar Skali Lami, Nikolaos Trichakis, Thomas Trikalinos, Mohammad Fazel Zarandi, Dimitris Bertsimas, Overview of DELPHI model V3 - COVIDAnalytics, 2020.

[23] Youyang Gu, COVID-19 projections using machine learning, 2020, https://covid19-projections.com.

[24] Seth Flaxman, et al., Estimating the effects of non-pharmaceutical interventions on COVID-19 in europe, Nature 584 (2020) 257–261.

[25] Seth Flaxman, et al. Swapnil Mishra, Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in European countries: technical description update, 2020, arXiv:2004.11342.

[26] European Centre for Disease Prevention and Control, Updated projections of COVID-19 in the EU/EEA and the UK, Technical report, ECDC: Stockholm, 2020.

[27] Lauren Castro1, Geoffrey Fairchild, Isaac Michaud, Dave Osthus, COFFEE: COVID-19 forecasts using fast evaluations and estimation, 2020, Los Alamos National Laboratory, LA-UR-20-28630.

[28] Ottavia Prunas, Joshua L Warren, Forrest W Crawford, Sivan Gazit, Tal Patalon, Daniel M Weinberger, Virginia E Pitzer, Vaccination with BNT162b2 reduces transmission of SARS-CoV-2 to household contacts in Israel, medRxiv : the preprint server for health sciences (2021) http://dx.doi.org/10.1101/2021.07.13.21260393.

[29] Jiwon Jung, Ji Yeun Kim, Heedo Park, Sunghee Park, Joon Seo Lim, So Yun Lim, Seongman Bae, Young Ju Lim, Eun Ok Kim, Jineui Kim, Man Seong Park, Sung Han Kim, Transmission and infectious SARS-CoV-2 shedding kinetics in vaccinated and unvaccinated individuals, JAMA Netw. Open 5 (2022) http://dx.doi.org/10.1001/JAMANETWORKOPEN.2022.13606, URL https://pubmed.ncbi.nlm.nih.gov/35608859/.

[30] Laith J. Abu-Raddad, Hiam Chemaitelly, Houssein H. Ayoub, Patrick Tang, Peter Coyle, Mohammad R. Hasan, Hadi M. Yassine, Fatiha M. Benslimane, Hebah A. Al-Khatib, Zaina Al-Kanaani, Einas Al-Kuwari, Andrew Jeremijenko, Anvar Hassan Kaleeckal, Ali Nizar Latif, Riyazuddin Mohammad Shaik, Hanan F. Abdul-Rahim, Gheyath K. Nasrallah, Mohamed Ghaith Al-Kuwari, Adeel A. Butt, Hamad Eid Al-Romaihi, Abdullatif Al-Khal, Mohametabd H. Al-Thani, Roberto Bertollini, Relative infectiousness of SARS-CoV-2 vaccine breakthrough infections, reinfections, and primary infections, Nature Commun. 13 (1) (2022) 532, http://dx.doi.org/10.1038/s41467-022-28199-7.

[31] Anika Singanayagam, Seran Hakki, Jake Dunning, Kieran J. Madon, Michael A. Crone, Aleksandra Koycheva, Nieves Derqui-Fernandez, Jack L. Barnett, Michael G. Whitfield, Robert Varro, Andre Charlett, Rhia Kundu, Joe Fenn, Jessica Cutajar, Valerie Quinn, Emily Conibear, Wendy Barclay, Paul S. Freemont, Graham P. Taylor, Shazaad Ahmad, Maria Zambon, Neil M. Ferguson, Ajit Lalvani, Anjna Badhan, Simon Dustan, Chitra Tejpal, Anjeli V. Ketkar, Janakan Sam Narean, Sarah Hammett, Eimear McDermott, Timesh Pillay, Hamish Houston, Constanta Luca, Jada Samuel, Samuel Bremang, Samuel Evetts, John Poh, Charlotte Anderson, David Jackson, Shahjahan Miah, Joanna Ellis, Angie Lackenby, Community transmission and viral load kinetics of the SARS-CoV-2 delta (b.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study, Lancet Infect. Dis. 22 (2022) 183–195, http://dx.doi.org/10.1016/S1473-3099(21)00648-4.

[32] Stephen M. Kissler, Joseph R. Fauver, Christina Mack, Caroline G. Tai, Mallery I. Breban, Anne E. Watkins, Radhika M. Samant, Deverick J. Anderson, Jessica Metti, Gaurav Khullar, Rachel Baits, Matthew MacKay, Daisy Salgado, Tim Baker, Joel T. Dudley, Christopher E. Mason, David D. Ho, Nathan D. Grubaugh, Yonatan H. Grad, Viral dynamics of SARS-CoV-2 variants in vaccinated and unvaccinated persons, New England J. Med. 385 (2021) 2489–2491, http://dx.doi.org/10.1056/NEJMC2102507, URL https://pubmed.ncbi.nlm.nih.gov/34941024/.

[33] Frederik Plesner Lyngse, Laust Hvas Mortensen, Matthew J. Denwood, Lasse Engbo Christiansen, Camilla Holten Møller, Robert Leo Skov, Katja Spiess, Anders Fomsgaard, Ria Lassaunière, Morten Rasmussen, Marc Stegger, Claus Nielsen, Raphael Niklaus Sieber, Arieh Sierra Cohen, Frederik Trier Møller, Maria Overvad, Kåre Mølbak, Tyra Grove Krause, Carsten Thure Kirkeby, Household transmission of the SARS-CoV-2 omicron variant in Denmark, Nat. Commun. 2022 13:1 13 (2022) 1–7, http://dx.doi.org/10.1038/s41467-022-33328-3, URL https://www.nature.com/articles/s41467-022-33328-3.

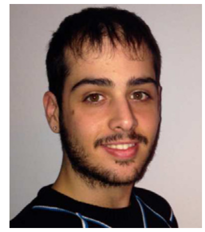
[34] Zachary J. Madewell, Yang Yang, Ira M. Longini, M. Elizabeth Halloran, Natalie E. Dean, Household secondary attack rates of SARS-CoV-2 by variant and vaccination status: An updated systematic review and meta-analysis, JAMA Netw. Open 5 (2022) E229317, http://dx.doi.org/10.1001/JAMANETWORKOPEN.2022.9317, URL https://pubmed.ncbi.nlm.nih.gov/35482308/.

[35] Julia M. Baker, Jasmine Y. Nakayama, Michelle O'Hegarty, Andrea McGowan, Richard A. Teran, Stephen M. Bart, Katie Mosack, Nicole Roberts, Brooke Campos, Alina Paegle, John McGee, Robert Herrera, Kayla English, Carla Barrios, Alexandria Davis, Christine Roloff, Lynn E. Sosa, Jessica Brockmeyer, Lindsey Page, Amy Bauer, Joshua J. Weiner, Manjeet Khubbar, Sanjib Bhattacharyya, Hannah L. Kirking, Jacqueline E. Tate, SARS-CoV-2 b.1.1.529 (omicron) variant transmission within households - four U.S. jurisdictions, november 2021-february 2022, MMWR. Morb. Mort. Wkly.

Rep. 71 (2022) 341–346, http://dx.doi.org/10.15585/MMWR.MM7109E1, URL https://pubmed.ncbi.nlm.nih.gov/35238860/.
[36] Sophia T. Tan, Ada T. Kwan, Isabel Rodríguez-Barraquer, Benjamin J. Singer, Hailey J. Park, Joseph A. Lewnard, David Sears, Nathan C. Lo, Infectiousness of SARS-CoV-2 breakthrough infections and reinfections during the omicron wave, Nat. Med. 29 (2) (2023/02/01) 358–365, http://dx.doi.org/10.1038/s41591-022-02138-x.
[37] Jason Mair, Zhiyi Huang, David Eyers, Yawen Chen, Quantifying the energy efficiency challenges of achieving exascale computing, in: Proceedings - 2015 IEEE/ACM 15th International Symposium on Cluster, Cloud, and Grid Computing, CCGrid 2015, Institute of Electrical and Electronics Engineers Inc., 2015, pp. 943–950, http://dx.doi.org/10.1109/CCGRID.2015.130.
[38] Bilge Acun, Kavitha Chandrasekar, Laxmikant V. Kale, Fine-Grained Energy Efficiency Using Per-Core DVFS with an Adaptive Runtime System, in: 2019 10th International Green and Sustainable Computing Conference, IGSC 2019, Institute of Electrical and Electronics Engineers Inc., 2019, http://dx.doi.org/10.1109/IGSC48788.2019.8957174.
[39] Pawel Bratek, Lukasz Szustak, Roman Wyrzykowski, Tomasz Olas, Tomasz Chmiel, Heterogeneous voltage frequency scaling of data-parallel applications for energy saving on homogeneous multicore platforms, in: Euro-Par 2021: Parallel Processing Workshops: Euro-Par 2021 International Workshops, Lisbon, Portugal, August 30-31, 2021, Revised Selected Papers, Springer-Verlag, Berlin, Heidelberg, 2021, pp. 141–153, http://dx.doi.org/10.1007/978-3-031-06156-1_12.
[40] Manjari Gupta, Lava Bhargava, Indu Sreedevi, Dynamic voltage frequency scaling in multi-core systems using adaptive regression model, in: Proceedings of the 4th International Conference on IoT in Social, Mobile, Analytics and Cloud, ISMAC 2020, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 1201–1206, http://dx.doi.org/10.1109/I-SMAC49090.2020.9243505.
[41] Péter Piros, Tamás Ferenci, Rita Fleiner, Péter Andréka, Hamido Fujita, László Főző, Levente Kovács, András Jánosi, Comparing machine learning and regression models for mortality prediction based on the hungarian myocardial infarction registry, Knowl.-Based Syst. 179 (2019) 1–7, http://dx.doi.org/10.1016/j.knosys.2019.04.027, URL https://www.sciencedirect.com/science/article/pii/S0950705119302060.
[42] Mahya Seyedan, Fereshteh Mafakheri, Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities, J. Big Data 7 (1) (2020/07/25) 53, http://dx.doi.org/10.1186/s40537-020-00329-2.

Miguel Guzmán is a researcher at the Universidad Carlos III de Madrid which he joined in 2020. He obtained a B.Sc. in Industrial Electronic in 2018 and an M.S. in Robotics Intelligence in 2020. His research interests include high- performance computing and computational epidemiology. He has been involved in the project "Medium and long-term forecast of COVID-19 propagation" funded by the Health Institute Carlos III and the project "Development of a tool for prediction of epidemiological scenarios and vaccination against COVID-19" funded by the Ministry of Health. He has authored several conference and journal papers in the field of computational epidemiology.

Maria-Cristina Marinescu has a PhD in Computer Science from UCSB. She has published more than 45 articles, was PI of 4 competitive projects, and has been involved in technology transfer. Her current research interests are semantic technologies, machine learning, and programming languages.



Alberto Cascajo has a Ph.D. in Computer Science and Technology at the Carlos III University of Madrid since 2021. I completed a Bachelor's Degree in Computer Science and a Master's Degree in Computer Science and Technology at the same university, being part of the ARCOS research group since 2013. Initially, my work focused on the development of railway construction simulators. As of 2017, I combine my studies and my thesis in the area of HPC and Big Data, with teaching in certain subjects within Computer Science.

My work focuses on the design of a monitoring system capable of offering an overview of large distributed systems, as well as identifying applications that are running to do scheduling based on different criteria (execution time, energy consumed, the maximum number of nodes, etc.). All of these, taking into account overheads, scalability, fault tolerance, etc. I am in the last stage of my doctorate.

Jesus Carretero is a full Computer Architecture professor at Universidad Carlos III de Madrid (Spain). His major research areas are parallel and distributed systems, focused on HPC and I/O systems optimization. He is PI of the EuroHOPC project ADMIRE and was coordinator of therNESUS research network on sustainable computing.

He is associate editor of IEEE transactions on Systems, Future Generation Computing Systems, COmputing Surveys. He has published more than Currently he coordinates the Informatics area Agency.

David E. Singh received the Ph.D. degree in Physics (Electronics and Computer Science) at the University of Santiago de Compostela in 2003. He is, since 2005, Associate Professor at the University Carlos III of Madrid, Spain, member of the Computer Architecture Group. His research interests include high-performance computing, automatic code optimization, parallel I/O, and computational epidemiology. He has authored more than 80 papers published in scientific journals as well as international conferences.