



# The CP-ABM approach for modelling COVID-19 infection dynamics and quantifying the effects of non-pharmaceutical interventions

Aleksandar Novakovic<sup>a,b,\*</sup>, Adele H. Marshall<sup>a,b,\*</sup>

<sup>a</sup> School of Mathematics and Physics, Queen's University Belfast, University Road, Belfast, BT7 1NN, Northern Ireland, United Kingdom

<sup>b</sup> Joint Research Centre in AI for Health and Wellness, Faculty of Business and IT, Ontario Tech University, 2000 Simcoe Street North, Oshawa, Ontario L1G 0C5, Canada

## ARTICLE INFO

### Article history:

Received 31 March 2021

Revised 4 May 2022

Accepted 11 May 2022

Available online 14 May 2022

### Keywords:

COVID-19

Non pharmaceutical interventions

Change point detection

Agent based model

Genetic algorithm

## ABSTRACT

The motivation for this research is to develop an approach that reliably captures the disease dynamics of COVID-19 for an entire population in order to identify the key events driving change in the epidemic through accurate estimation of daily COVID-19 cases. This has been achieved through the new CP-ABM approach which uniquely incorporates Change Point detection into an Agent Based Model taking advantage of genetic algorithms for calibration and an efficient infection centric procedure for computational efficiency.

The CP-ABM is applied to the Northern Ireland population where it successfully captures patterns in COVID-19 infection dynamics over both waves of the pandemic and quantifies the significant effects of non-pharmaceutical interventions (NPI) on a national level for lockdowns and mask wearing.

To our knowledge, there is no other approach to date that has captured NPI effectiveness and infection spreading dynamics for both waves of the COVID-19 pandemic for an entire country population.

Crown Copyright © 2022 Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## 1. Introduction

On the 30<sup>th</sup> January 2020, the World Health Organization declared a global outbreak, and on the 11<sup>th</sup> March of the same year, it declared the pandemic of coronavirus disease 2019 (COVID-19). COVID-19 is caused by a respiratory virus named SARS-CoV-2, with its main mode of human-to-human transmission via direct, indirect or close contact with an infected person through their infected secretion (droplets, aerosols, saliva) which is produced when an infected person sneezes, coughs, talks or sings [1].

One of the main characteristics of this virus is that the people who get infected do not develop any symptoms immediately upon its contraction (i.e. become **symptomatic** or **clinical**). The period of time that passes between exposure to the virus and symptom onset is called the **incubation period** [1]. This incubation period varies from person to person, lasting for a median of 5.1 days (95% CI = [4.5, 5.8]), but in some instances its duration can be as long as 14 days [2]. However studies have also demonstrated that infected people without symptoms, that is those who are in the in-

cubation period and not yet symptomatic (i.e. **presymptomatic** or **preclinical** careers) and those who never develop any symptoms (i.e. **asymptomatic** or **subclinical** careers), can spread the disease too [3]. Given that the subclinical people are not always tested, numerous research findings have suggested that the total number of infections is very likely greater than the number of reported cases [4]. The true proportion of COVID-19 transmissions that is accounted for by preclinical or subclinical people is unknown and as such this has big implications for prevention [5].

To fight the global pandemic, the nations around the world have been introducing a range of different non-pharmaceutical interventions (NPIs) that aim to control and reduce the rapid spread of the virus. These NPIs include, but are not limited to, partial and/or total lockdowns on both a regional and/or national level, promoting social distancing and the making wearing of face masks mandatory on public transport, indoors and/or outdoors, etc. Although these measures are proven to be effective in reducing the spread of disease [6], their enforcement affects sociopolitical, economical and all the other aspects of life. Given the complexity of societies and differences in the interventions that different nations are taking in fighting COVID-19, it is very difficult to predict their short and medium term impact both on a national and global level [7,8]. Some interventions such as regional and total national lock-

\* Corresponding authors, joint first authors.

E-mail addresses: [a.novakovic@qub.ac.uk](mailto:a.novakovic@qub.ac.uk) (A. Novakovic), [a.h.marshall@qub.ac.uk](mailto:a.h.marshall@qub.ac.uk) (A.H. Marshall).

downs, can have a devastating impact on the national economies and mental health of society particularly those most vulnerable which is why it is particularly important to quantify the effect that these NPIs have on the reduction of the spread of the virus, so that taking any similar approaches can be justified in future decisions.

### 1.1. Related work

When it comes to modelling the spread of COVID-19 the three approaches most commonly found in the literature are the compartmental, AI and agent-based modelling approaches, each with their own advantages and disadvantages.

#### 1.1.1. Compartmental and AI modelling approaches

Compartmental models belong to the equation based models group, and provide a theoretical framework for describing disease dynamics and analysing a specific outbreak or epidemic within a closed and well mixed homogenous population [9,10]. Each compartment represents one disease status, and each individual in the population can be in exactly one compartment in a given time but can move from one compartment to another depending on the model parameters [11].

A compartmental model consists of a system of differential equations, with each differential equation representing a single compartment in the model [12]. There are many types of these models, but given that there is a known incubation period for COVID-19, the SEIR (Susceptible-Exposed-Infected-Recovered) compartmental models [13] are most frequently used for modelling its disease dynamics as for instance the research by Kuniya [14]. By adding new compartments, many refinements of the standard SEIR COVID-19 models have been made in order to create more realistic models that for example include: super spreaders [15], preclinical and subclinical [16], quarantined [17], and other types of patient.

Compartmental models represent the majority among those that can be found in the literature for simulating COVID-19 disease dynamics [8]. Their main advantage is in their capability to capture large scale infection dynamics at a macro level (e.g. country-wide/continent-wide pandemic), with relatively low computational overhead. However, at the same time this top down approach is usually being listed as their biggest limitation in the literature, as they are unable to capture more refined information on the spread of disease such as for instance the interaction between individuals [10].

More recently with the increasing data availability, AI approaches have been gaining more traction within COVID-19 modelling in particular, for diagnostic purposes from medical images [18–20] and for forecasting the spread of the disease. An example of the former is, Wang et al. [18] who proposed a new framework that utilises deep learning to differentiate and localise COVID-19 from chest X-ray images of community acquired pneumonia patients. Similarly, an example of the latter case is the work of ArunKumar et al. [21] who use deep layer Recurrent Neural Networks (RNNs) with Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) cells to estimate 60 day forecasts of COVID-19 cases for the top ten countries. However, the AI models require a large amount of training data which consume an exhaustive amount of training time. When it comes to modelling infection dynamics, the focus of this paper, the AI models have similar limitations as the compartmental models with being a top down approach. This includes an inability to capture more refined information on the spread of disease such as for instance the interaction between individuals, the change in their attitudes in response to the NPIs, the effect that this may have on the spread dynamics and the difficulty in quantifying the NPI effectiveness [22].

#### 1.1.2. Agent based modelling approaches

Agent Based Models (ABMs), take the opposite approach to the macro level view of the compartmental and AI models, by creating a micro level view and simulating it up to the macro level. The ABMs are computer based simulations that consist of heterogeneous and adaptive individual entities called **agents** that are uniquely identifiable, and capable of acting autonomously and interacting with one another in the simulated environment [23,24]. Their main characteristic is that by defining the set of micro-rules that describe the agents' behaviour in the simulated environment, they are able to capture emergent macro effects and realistically model a real world system [25]. More precisely, when it comes to epidemiological modelling, the behaviour of the agents combined with the transmission pattern and disease progression will lead to the emerging population dynamics such as a disease outbreak or pandemic [24]. In order to be as realistic as possible so that the results of the ABMs can be applied on a population level, the characteristics of agents, their behaviour and the characteristics of the disease in the simulated environment should be as close as possible to the ones that can be found in the real world. However this capability of ABMs to mimic real world scenarios comes with a price, as with the increased level of details that are captured by the models, comes a requirement for increased computing power to run these simulations. Therefore, the scalability related issue when creating highly detailed models, is often listed as one of the main limitations of ABMs as noted in the literature [10].

The majority of the previously published ABMs used for modelling the spread of COVID-19 have focused on a very small time interval almost all of which covering the first wave only. The limiting factor is likely to have been due to the big issue of scalability as national level models need to have the number of agents in the model matching the number of individuals in the population (based on their census). It is not surprising therefore that no papers can be found reporting the development of an ABM at a national level for both the first and second waves and only two that fit the first wave of the pandemic for their entire population [26,27].

Chang et al. [26] use an ABM to model the best COVID-19 strategies for intervention in Australia, such as restricting international travel, school closures and social distancing. The AceMod, an established Australian-census ABM, was calibrated by using real data until the end of June 2020 for the age-dependent fraction of symptomatic cases and simulations conducted for the period of March 2020 to August 2020. The model runs on a high performance cluster with 4264 cores with over 24 million software agents which mimic the population of Australia.

Hinch et al. [27] developed an ABM for evaluating the impact of NPIs in the UK population based on the UK 2011 Census data for household size and age structure of the population of 65 million. The OpenABM-Covid19 model simulates the UK population from the date of the first national lockdown on the 23<sup>rd</sup> March 2020 until 60 days later. The authors report on the model realistically simulating an outbreak, the corresponding hospitalization and ICU admissions for a default population of 1 million inhabitants in an urban environment where each day takes approximately 3 seconds to run and requires 5Gb of memory (reduced to 1.7Gb if contact-tracing is disabled) on a 2015 MacBook Pro. The model has been developed using the C language but is run via Python using a SWIG-interface. A framework of over 200 tests are used to validate the model. There is no detail provided on how the validation is performed except to report that the expected outputs match the specified set of input parameters.

More commonly, due to high computational costs, researchers using ABMs typically focus on modelling the COVID-19 spread at a regional or city level or by using a smaller number of agents than

the entire population and then extrapolate the results. Examples of such are as follows:

1. Urmia city, Iran is represented by an ABM with 750,805 agents for each individual of the population based on the Census 2017 data [28]. A chi-square test was performed to test whether the average number of weekly infected agents for 100 runs of the ABM were significantly different to the actual observed data. The results indicated that this was not significant and hence the model was considered a suitable representation of Urmia city. No further calibration was performed.
2. In the Chernivtsi region of Ukraine, an ABM was developed to predict the spread of COVID-19 using 1000 agents and further applied to regions in Slovakia, Turkey and Serbia using between 500–1000 agents [29]. The authors include visual comparisons of the real data and forecasted spread to demonstrate the approach's suitability and report that statistical results, and sensitivity analysis were also conducted. The method is noted to be slower and disadvantaged due to the dependence on the random number generator which can produce different simulation results for the same initial parameters. This is overcome by parallelizing the models on different cores/computers.
3. In Italy, the Calabria region was modelled using an ABM with a closed population of 250 agents moving within a square section of  $250 \times 250 \text{ m}^2$  [9]. The number of agents was kept low to minimize the computational cost, which was reported to have an average CPU time for the simulations over the 90 day period to be 3.5 h on a computer with four cores and 6Gb of RAM. The model was assessed visually by comparing the simulated number of cases with those reported for Calabria and other regions of Italy.
4. In Brazil, an ABM was created to replicate a closed society consisting of 300 agents made up of 5 agent types (people, houses, businesses, the government and the healthcare system) and simulated for a 2 month period [8]. The paper also models the financial impact, however there is no information provided on the calibration of the model.
5. Across Ireland an ABM previously developed for modelling the spread of measles was extended into a hybrid model to represent COVID-19 infection for 12 small towns, with one agent representing each individual in the populations ranging in size from 73 to the largest town at 1782 people [10]. The overall hybrid model was compared with the ABM but no calibration with the real data was discussed.
6. In France, an ABM was constructed for 500,000 agents and run for a period of 360 days [30]. The results were then extrapolated for the entire 64 million population of France. The model contained 194 parameters, 192 of which were estimated from the literature with the remaining two calibrated using the Nelder Mead method. To the best of our knowledge, this paper is the only one that fully calibrates their model using a number of different methods from visual inspection of the fit of the predicted to the observed data for the ICU admissions, bed occupancy and mortality (both daily and cumulative) to an evaluation of the resulting  $R^2$  and Pearson's R estimates, ranging between 0.87 and 0.93. The Cochran–Mantel–Haenszel test was also carried out to compare the observed and predicted age distribution of deceased people with the conclusion that there was no significant difference.

Some other examples of developing ABMs for COVID-19 applications consider the model of spread in facilities such as the one developed by Cuevas [31] who conduct a number of experiments to mimic different scenarios in a facility where there is between 10 and 1000 agents, and the one developed by Araya [32] who model 100 construction worker agents working on one specific project every hour of their working day for a period of 3 working months.

Due to the lack of data, both models are hypothetical in nature and based on subject matter expert information rather than data.

## 1.2. Research motivation

The model presented in this paper is motivated by challenges faced by the authors when modelling the infection spread of COVID-19 for the Northern Ireland COVID-19 Modelling Group. Previous models listed a large number of fixed parameters which requires making a large number of assumptions which may be unrealistic or are too specific to the population in question to be easily transferable to Northern Ireland. There is only one paper [30] that clearly describes the calibration performed on their model and in many no description or calibration appears to be performed. Likewise, for those papers that consider a subsection of their population and wish to expand that to the entire population, it is unclear how extrapolation is performed if any. Additionally, all of the models previously reported are considering only wave one of COVID-19 pandemic and run the models for a short period of time. The main contributions of this paper are as follows:

1. We propose a novel methodology based on the hybridisation of change point detection and agent based modelling techniques for modelling COVID-19 infection dynamics, and quantifying the effects of non-pharmaceutical interventions on a national level. The proposed methodology is programming language agnostic and enables researchers to develop models that can be run and calibrated on consumer grade hardware and not necessarily just on supercomputers.
2. We demonstrate effectiveness of the methodology introduced by modelling the spread of the new daily confirmed COVID-19 cases in Northern Ireland during the period between the 9<sup>th</sup> March and 15<sup>th</sup> November 2020. This is accompanied by a detailed description of the entire calibration and validation process, as well as interpretations of the fitted parameters.
3. We are able to successfully capture the role that subclinical (also known as the asymptomatic) patients have on the spread of the virus and estimate their relative infectiousness, as well as successfully quantify the NPI effects that the wearing of masks, one regional and two national lockdowns have had on the virus spread reduction.
4. As far as we are aware of, this study is the first of its kind that was able to successfully capture COVID-19 infection dynamic during such a long period of time and successfully model both infection waves.

The paper proceeds as follows. In [Section 2](#) we introduce the CP-ABM methodology and provide detailed descriptions of the algorithms of each of its components. In [Section 3](#) we describe how we applied this methodology in Northern Ireland (as our case study) and provide a thorough discussion of the obtained results making parallels with the outcomes that can be found in other independent research studies and literature. Finally, in [Section 4](#) we give our closing remarks and lay out directions for our future research.

## 2. Methodology

The main objective of the CP-ABM modelling approach is to realistically simulate the COVID-19 infection transmission within a closed society living in a shared environment during the observed time period. The model assumes that a disease transmission may occur only when the infectious (preclinical, clinical or subclinical) person gets in direct contact with a susceptible person, and that changes in the contact behaviour of people is the driving force behind COVID-19 infection rates [33].

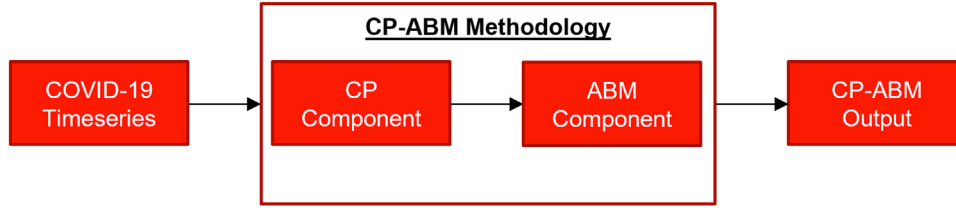


Fig. 1. A schematic overview of the CP-ABM methodology.

The CP-ABM model consists of two components: the change point detection (CP) component and the agent based modelling component (ABM). The **change point detection (CP) component** is used to identify key events that result in contact behaviour changes (i.e. mean daily contacts changes) of individuals living in the modelled society and which affect the disease transmission the most during the observed time period. The **agent-based modelling (ABM) component** takes the outputs from the CP component and adds other NPIs that do not directly influence the mean daily contacts but that have an influence on infection dynamics (such as the mandatory wearing of masks) to appropriately adjust the behaviour of its agents over the course of the simulation and produce realistic results (Fig. 1).

### 2.1. An overview of the change point detection (CP) component

Many timeseries datasets can exhibit a sudden change in structure, such as unexpected jumps or drops in level or volatility [34]. The indexes that denote the moments in time in which the characteristics of timeseries abruptly change are called **change points**. Change points usually indicate changes in the underlying data generation mechanism, and given that the changes in social mixing patterns have a detrimental impact on the COVID-19 spreading dynamics, in our work we associate change points with the moments in the timeseries that correspond to those non-pharmaceutical interventions (NPIs) that played a critical role in governing contact behaviour between individuals in the population.

In order to explain how the CP component works we need to introduce some basic definitions. Let us assume that the observed data for the confirmed daily COVID-19 cases is represented by  $\mathcal{D} = \langle d_k \rangle_{k=1}^N$ , where  $N \geq 1$ , is a sequence of  $N$  observations in  $\mathbb{N}$ . Also, let  $\mathcal{E} = \langle e_{(k)} \rangle_{k=1}^Q$  be an increasing sequence of  $Q$  integers that can take the values between 1 and  $N$  (inclusive), representing the indexes of occurrence of all NPIs that were used for regulating social-mixing patterns in  $\mathcal{D}$  (e.g. lockdowns, opening/closing schools, etc.). The main task of the CP component is to find  $\mathcal{T} \subset \mathcal{E} = \langle t_{(k)} \rangle_{k=1}^M$ , an increasing sequence of  $M$  integers that can take the values between 1 and  $N-1$  (inclusive), that represent the change points in  $\mathcal{D}$  and correspond to the key NPIs from  $\mathcal{E}$  that were most influential on the spread of the disease.

These change points  $\mathcal{T}$  segregate the timeseries  $\mathcal{D}$  into  $M+1$  non-overlapping segments  $\langle d_{(t_{k-1}, t_k]} = \langle d_{t_{k-1}+1}, \dots, d_{t_k} \rangle \rangle_{k=1}^{M+1}$ , where  $t_0 = 0$  and  $t_{M+1} = N$ . In practice the exact number of change points is usually unknown [34], hence the problem of estimating  $\mathcal{T}$  can be formulated as a model selection problem where the objective is to find the best segmentation  $\hat{\mathcal{T}}$  that minimizes the following quantification criteria:

$$\min_{M, \hat{\mathcal{T}}} \left\{ \sum_{k=1}^{M+1} \mathcal{C}(d_{(\hat{t}_{k-1}, \hat{t}_k]}) + f(M) \right\} \quad (1)$$

where  $\mathcal{C}(\cdot)$  represents a cost function that measures a goodness of fit for segment  $d_{(\hat{t}_{k-1}, \hat{t}_k]}$ , while  $f(M)$  is a carefully selected penalty parameter that increases with  $M$  [35]. This penalty parameter is used to prevent overfitting by controlling the number of change points  $M$  and thus penalizing for complexity.

The choice of  $\mathcal{C}(\cdot)$  most often depends on the assumptions of how the distribution of the underlying dataset  $\mathcal{D}$  is parametrised. When it comes to the spread of COVID-19, a growing body of evidence suggests that the number of confirmed cases follows a power-law distribution [36]. Fagan et al. [37] demonstrate that no single change point detection method fully captures the behaviour of the heavy-tailed nature of the data of power-law distributions. To mitigate these issues, they propose a hybrid solution where the optimal segmentation that minimizes (1) is found by utilising the ED-PELT non-parametric change point detection algorithm in which the empirical distribution of  $\mathcal{D}$  is used to define  $\mathcal{C}(\cdot)$  [38], in conjunction with two penalty choices for  $f(M)$ : the modified Bayesian information criterion (mBIC) [39] and the change points for a range of penalties (CROPS) [35].

A schematic overview of the CP component that we use is shown in Fig. 2. The approach of Fagan et al. [37] is utilised in the approximation phase of our algorithm for the CP Detection Component to approximate the locations where the statistical properties of  $\mathcal{D}$  change. The pseudo-code of the approximation phase along with the identification, confirmation and segmentation phases is provided in Algorithm 1. The main task of the approximation phase (lines 2-9) is to find the optimal number of change points and their location in  $\mathcal{D}$ , and in the case of misalignment to adjust the positions of detected change points to match the nearest NPIs from  $\mathcal{E}$  that were crucial in influencing social-mixing patterns in the population (*identification phase*, lines 10-19). This component also enables the modeller to include any additional NPIs of interest that are not fully captured by the previous two phases (*NPI confirmation phase*, lines 20-24) and to produce the final segmentation  $\mathcal{T}$  (*segmentation phase*, line 25) to inform the ABM component of the model to appropriately adapt the behaviour of its agents during the simulation.

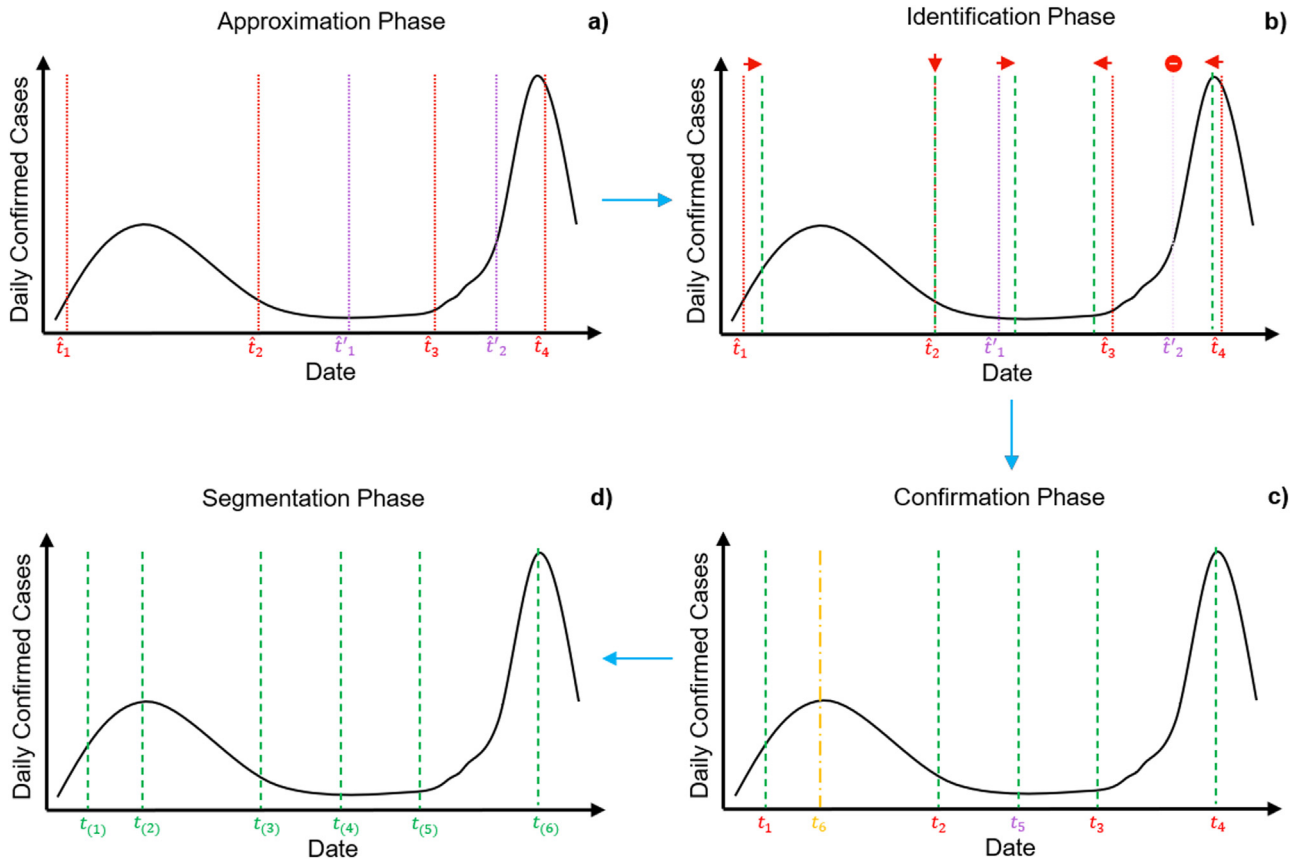
### 2.2. An overview of the agent based-model component

The main objective of the ABM component is to realistically simulate COVID-19 infection dynamics inside the closed heterogeneous population that is inhabiting the shared closed environment. It also aims to estimate the effectiveness of the key NPIs that lead to the changes in population mixing dynamics, as well as the effectiveness that mask wearing, has had on the disease spread during the observation time period. This ABM component consists of two types of entities: **agents** representing the age-stratified individuals inside the simulated population and **cells** representing the environment in which these agents exist and interact. There are 7 cells in total  $\mathcal{C} = \{S^s, S, E, I^s, I^p, I^c, R\}$ , with each cell  $\zeta \in \mathcal{C}$  describing one of the mutually exclusive infection status:  $S^s$  (shielded susceptible),  $S$  (susceptible),  $E$  (exposed),  $I^s$  (infectious and subclinical),  $I^p$  (infectious and preclinical),  $I^c$  (infectious and clinical) and  $R$  (recovered). Each cell can contain multiple agents, with agents inheriting the infection status of the cell in which they are located (e.g. all agents that are on the cell with infection status  $S$  are susceptible, while those on the cell with status  $E$  are exposed, etc.). Let us denote with  $\mathcal{A}^{\mathcal{C}} = \{\{a_k^{\zeta}\}_{k=1}^{N_A} | N_A \in \mathbb{N}^+, \zeta \in \{S^s, S, E, I^s, I^p, I^c, R\}\}$  a set of  $N_A \in \mathbb{N}^+$  agents that represent all individuals in the simu-



**Table 1**  
Overview of the entities and their attributes of the ABM component.

| Entity | Attribute      | Symbol         | Possible Values  | Description  |
|--------|----------------|----------------|--|--|
| cell   | Status         | $\zeta$        | $S^S, S, E, I^S, I^P, I^C, R$  | Infection status representation  |
| agent  | n-agents       | $\eta^\zeta$   | $\mathbb{N}^+$   | Number of agents residing at the cell  |
|        | Age            | $g$            | $[0, 90]$  | Age in years   |
|        | Status         | $\zeta$        | $S^S, S, E, I^S, I^P, I^C, R$  | Agent's infection status inherited from the status of the cell in which it is located  |
|        | Susceptibility | $s^g$          | $= \begin{cases} 0.008, & \text{if } g \in [0, 9] \\ 0.007, & \text{if } g \in [10, 19] \\ 0.04, & \text{if } g \in [20, 29] \\ 0.06, & \text{if } g \in [30, 39] \\ 0.08, & \text{if } g \in [40, 49] \\ 0.16, & \text{if } g \in [50, 59] \\ 0.38, & \text{if } g \in [60, 69] \\ 0.52, & \text{if } g \in [70, 90] \end{cases}$ | Age stratified susceptibility to infection [40]  |
|        | Duration       | $\delta^\zeta$ | $\sim \begin{cases} [\Gamma(\mu = 3.0, k = 4)], & \text{if } \zeta = "E" \\ [\Gamma(\mu = 2.1, k = 4)], & \text{if } \zeta = "I^P" \\ [\Gamma(\mu = 2.9, k = 4)], & \text{if } \zeta = "I^C" \\ [\Gamma(\mu = 5.0, k = 4)], & \text{if } \zeta = "I^S" \end{cases}$  | Duration in days that an agent spends in the $E, I^S, I^P, I^C$ cells before moving to another cell in the infection dynamics workflow (Fig. 3) [16] |



**Fig. 2.** A schematic overview of the processes inside the Change Point Detection Component.

lation, with the  $k^{\text{th}}$  agent belonging to one out of seven given cells  $\zeta \in \mathcal{C}$ . An overview of the attributes characterising cells and agents is provided in Table 1.

The ABM component conducts an iterative process (i.e. simulation) that is implemented through a series of discrete steps that we call **ticks**, with each tick ( $\tau$ ) in the simulation representing one day in the real world. The simulation starts by initialising the input parameters for the baseline relative infectiousness for each agent type, the number of agents at each state at the beginning of the simulation, the mean daily contacts in each segment between change points, the probability of the exposed agents developing a subclinical infection at the end of the incubation pe-

riod, and the variations attributed to additional behavioural effects. The output of the model is the estimated number of daily cases  $\hat{D} = \langle \hat{d}_k \rangle_{k=1}^N$  which is then used to evaluate the effectiveness of the NPIs by comparing the estimated values from a number of different scenarios against the true observed data  $D$ . The model is also able to implicitly capture the number of agents on each cell and track their interactions and behaviours at each tick, but for the focus of this paper we keep our attention on the number of estimated daily cases.

In order to produce outputs that are as realistic as possible, the number of agents to be used in the simulation and their age stratification should be initialised to match publicly available de-

**Algorithm 1**

The pseudo-code of the Change Point Detection (CP) Component of the CP-ABM model.

CP-COMPONENT( $\mathcal{D}, \mathcal{E}$ )**Input:**

$\mathcal{D} = \{\langle d_k \rangle_{k=1}^N \mid N \in \mathbb{N}^+\} \leftarrow$  A sequence representing a timeseries dataset containing the daily numbers of confirmed COVID-19 cases;  
 $\mathcal{E} = \{\langle e_k \rangle_{k=1}^Q \mid Q \leq N\} \leftarrow$  An increasing sequence of integers taking values between 1 and  $N$ , representing all indexes in  $\mathcal{D}$  that correspond to NPI that affected social mixing patterns

**Output:**

$\mathcal{T} = \langle t_{(k)} \rangle_{k=1}^M \leftarrow$  an ordered sequence of change points representing the key NPIs from  $\mathcal{E}$  that affected disease spreading dynamics the most

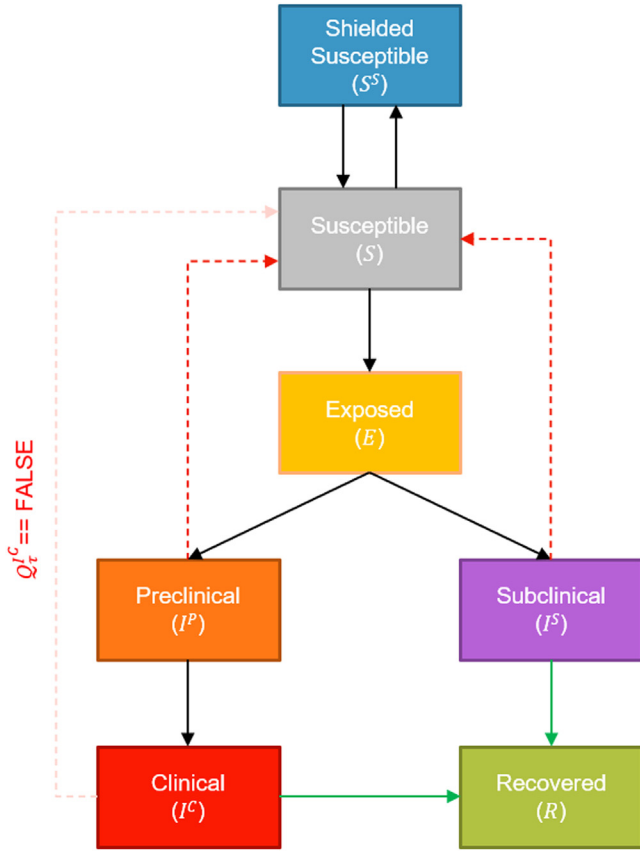
|   |   |
|---|---|
| <ol style="list-style-type: none"> <li>1. <math>\mathcal{T} \leftarrow \langle \rangle</math>; <math>k \leftarrow 1</math></li> <li>2. <math>\hat{\mathcal{T}}_{mBIC} \leftarrow</math> estimated set of change points by applying ED-PELT with mBIC penalty (Eq. 1)</li> <li>3. <math>\hat{\mathcal{T}}_{CROPS} \leftarrow</math> estimated set of change points by applying ED-PELT with CROPS (Eq. 1)</li> <li>4. <b>if</b> <math> \hat{\mathcal{T}}_{mBIC}  = 0</math> and <math> \hat{\mathcal{T}}_{CROPS}  &gt; 2</math> <b>then</b></li> <li>5.     <b>return</b> <math>\mathcal{T}</math></li> <li>6. <b>else</b></li> <li>7.     <math>\hat{\mathcal{T}} \leftarrow \hat{\mathcal{T}}_{mBIC} \cap \hat{\mathcal{T}}_{CROPS}</math></li> <li>8.     <math>\hat{\mathcal{T}}' \leftarrow (\hat{\mathcal{T}}_{mBIC} \cup \hat{\mathcal{T}}_{CROPS}) \setminus \hat{\mathcal{T}}</math> <math>\triangleright</math> interpretation required (steps 14–20)</li> <li>9.     <b>end if</b></li> <li>10.     <b>for each</b> <math>\hat{t} \in \hat{\mathcal{T}}</math> <b>do</b></li> <li>11.         <math>t \leftarrow \arg \min_{e \in \mathcal{E}}  \hat{t} - e </math></li> <li>12.         <math>\mathcal{T}_k \leftarrow t</math>; <math>k \leftarrow k + 1</math></li> <li>13.     <b>end for</b></li> <li>14.     <b>for each</b> <math>\hat{t}' \in \hat{\mathcal{T}}'</math> <b>do</b></li> <li>15.         <math>t \leftarrow \arg \min_{e \in \mathcal{E}}  \hat{t}' - e </math></li> <li>16.         <b>if</b> interpretation leads to conclusion that <math>t</math> is relevant <b>then</b></li> <li>17.             <math>\mathcal{T}_k \leftarrow t</math>; <math>k \leftarrow k + 1</math></li> <li>18.         <b>end if</b></li> <li>19.     <b>end for</b></li> <li>20.     <b>for each</b> <math>e \in \mathcal{E} \setminus \mathcal{T}</math> <b>do</b></li> <li>21.         <b>if</b> interpretation leads to conclusion that <math>e</math> is relevant <b>then</b></li> <li>22.             <math>\mathcal{T}_k \leftarrow e</math>; <math>k \leftarrow k + 1</math></li> <li>23.         <b>end if</b></li> <li>24.     <b>end for</b></li> <li>25.     <b>return</b> <math>\text{sort}(\mathcal{T})</math></li> </ol> | <p>Initialisation<br/> Approximation<br/> Phase (Fig. 2 a))<br/> [37]</p> <p>Identification<br/> Phase<br/> (Fig. 2 b))</p> <p>Confirmation Phase<br/> (Fig. 2 c))</p> <p>Segmentation<br/> Phase (Fig. 2 d))</p> |
|---|---|

mographics data of the modelled population or geographical region. It is important to highlight that the size of the agent set remains constant within the simulated time period, that is, there is no change in the population demographics. The number of agents to be randomly allocated to each cell  $\zeta \in \mathcal{C}$  during the initialisation phase ( $\tau = 0$ ), is controlled by the  $\eta_0^S$  parameter. In reality, the true number of agents with exposed, preclinical, subclinical and recovered infection status is unknown so the parameters  $\eta_0^S$ , where  $\zeta \in \{E, I^P, I^S, R\}$  need to be either assumed or estimated through calibration. In addition, there is the possibility that some susceptible agents are being protected and shielded from the rest of the other types of agent. Such individuals cannot be infected and are described as shielded susceptible. Shielded agents are located at the cell with status  $S^S$  and the time in which the shielding will occur will depend on the scenario being modelled.

It has been assumed that the agents' social mixing patterns are the driving force behind COVID-19 infection spread and that these patterns are directly affected by the key NPIs that are captured by the CP component at the change points  $\mathcal{T} = \langle t_{(k)} \rangle_{k=1}^M$ . These change points segregate the simulation time period into  $M + 1$  non-overlapping segments  $\langle (t_{k-1}, t_k] \rangle_{k=1}^{M+1}$ , with the agents' social mixing behaviour at tick  $\tau \in (t_{k-1}, t_k]$  being controlled by the mean daily contacts rate  $m_\tau = m_{(t_{k-1}, t_k]}$ , where  $k = 1, \dots, M + 1$ ;  $t_0 = 0$  and  $t_{M+1} = N$ . The mean daily contact rate  $\langle m_{(t_{k-1}, t_k]} \rangle_{k=1}^{M+1}$  on these segments are usually not known in advance, and therefore they need to be estimated via calibration.

Bearing in mind the modelling assumption that the infection can only spread throughout the population when there is an effective contact between infectious and susceptible individuals, the high computational costs can be reduced in the ABM component

by employing a novel infection centric approach which aims to capture the interactions of infectious agents only, i.e. the interactions of only those agents that are able to transmit a disease to others during the simulated time period. For every tick  $\tau$ , each infectious agent  $a^{\zeta^I}$  encounters a number ( $N_\tau^{a^{\zeta^I}}$ ) of other agents, where  $N_\tau^{a^{\zeta^I}}$  is drawn from a Poisson distribution with mean  $m_\tau$  [41]. Disease transmission from  $a^{\zeta^I}$  to each encountered agent  $a^{\zeta'}$  may happen only if the  $a^{\zeta'}$  is susceptible (and not shielding), and if a randomly drawn number from the uniform distribution  $\mathcal{U}(0, 1)$  is lesser than the probability of infection  $P_a^{S \rightarrow E}$  of the agent  $a^{\zeta'}$ . In the case of successful disease transmission, the agent  $a^{\zeta'}$  becomes exposed and moves from its cell with status  $S$  to the cell with status  $E$  for the period of  $\delta^E$  ticks, during which it is not infectious. At the end of the incubation period  $\delta^E$ , this agent will develop a subclinical infection with probability  $P^{E \rightarrow I^S}$  and move to cell  $I^S$  for the period of  $\delta^{I^S}$  ticks, after which it will recover (have status recovered). Otherwise, the agent will develop a preclinical infection and move to cell  $I^P$ . After spending a period of  $\delta^{I^P}$  ticks in the  $I^P$  cell, the agent will develop a clinical infection and move to cell  $I^C$  where it will stay until recovering, for a duration of  $\delta^{I^C}$  ticks. The recovery of the agent (either from subclinical or clinical infection) means that it is no longer infectious nor prone to reinfection, and requires its movement to the cell with infection status  $R$  until the end of simulation time period. The durations of time  $\delta^E, \delta^{I^S}, \delta^{I^P}$  and  $\delta^{I^C}$  (measured in ticks) represents the time that each infected agent spends in the cells  $E, I^S, I^P$  and  $I^C$ , respectively, and is approximated separately for each agent by rounding a randomly drawn number from the gamma distribution  $\Gamma(\mu, k)$  to the nearest integer, with  $\mu$  and  $k$  parameters provided in Table 1. The entire infection dynamics workflow is depicted in Fig. 3.



**Fig. 3.** An overview of the infection dynamics workflow where the solid arrows represent possible transitions of agents between the cells and the dashed arrows indicate interactions between the agents where the transition of infection may occur.

We always consider infectious agents to be all those agents with subclinical or preclinical infection, i.e. all agents that are located at the cells  $I^S$  and  $I^P$  at tick  $\tau$ . Given that the NPIs usually require isolation of symptomatic patients, we introduced the parameter  $Q_t^C$  that controls whether the agents with developed clinical infection are also going to be involved in spreading infection at tick  $\tau$  or not. Depending on the parameter  $Q_t^C$ , let  $\mathcal{A}_\tau^I = \{A_\tau^I | \zeta^I \in \mathcal{C}_\tau^I\}$  denote the set of all infectious agents located at the infectious cells  $\mathcal{C}_\tau^I = \begin{cases} \{I^S, I^P\}, & \text{if } Q_t^C = \text{TRUE} \\ \{I^S, I^P, I^C\}, & \text{if } Q_t^C = \text{FALSE} \end{cases}$  during tick  $\tau$ .

Likewise, we introduced the parameter  $Q_t^C$  that controls whether the agents with a developed clinical infection are also going to be involved in spreading infection at tick  $\tau$  or not.

Overall, the relative infectiousness of  $\mathcal{A}_\tau^I$  is affected by two main factors. The first factor is the baseline relative infectiousness  $\tau_b^{\zeta^I}$  of the agents  $A_\tau^I$  that depends on the cell  $\zeta^I \in \mathcal{C}_\tau^I$  in which they are located. We assume that the baseline relative infectiousness of agents with preclinical and clinical infection is 1 (corresponding to 100%) and that the agents who develop subclinical infection are less infectious than those who develop preclinical and clinical infection as suggested by the evidence from the literature [5,16]. Based on the fact that, there is no clear agreement in the literature about how much smaller their relative infectiousness is, the  $\tau_b^{\zeta^I}$  value needs to be obtained via calibration. Once initialised, the baseline relative infectiousness will remain constant throughout the simulation for all infectious agents.

The second factor that has a significant impact on the overall relative infectiousness of the agents  $\mathcal{A}_\tau^I$  are introductions and relaxations of non-pharmaceutical interventions such as social dis-

tancing, mask wearing enforcement or hygiene promoting rules, that cannot be solely captured by the decrease or increase of the mean daily contacts  $m_\tau$ . To intrinsically capture this extra variation that is attributed to these external behavioural effects, we introduce a behavioural parameter  $b_t$  that can either take a value of 0 if there are no non-pharmaceutical intervention taking place at tick  $\tau$  or otherwise take a real value from the  $[0,1]$  interval that is obtained again by calibration. The value of the  $b_t$  parameter varies from intervention to intervention, and requires separate fitting for each intervention.

The probability that the susceptible agent  $a^S$  will become infected at tick  $\tau$  depends on its age stratified susceptibility to infection  $s_a^g$  (Table 1) and on the overall relative infectiousness of the infectious agents  $a^I \in \mathcal{A}_\tau^I$  that it encountered, and thus calculated as  $P_a^{S \rightarrow E} = \tau_b^{\zeta^I} \times (1 - b_t) \times s_a^g$ . Pseudo-code describing the entire process of the COVID-19 infection spread at each tick  $\tau$  is provided in Algorithm 2. The algorithm starts by initialising cells with the infectious agents  $\mathcal{C}_\tau^I$  and the cells  $\mathcal{C}_\tau^I$  containing the agents that these infectious agents can contact at tick  $\tau$ . The only contact that cannot be made at any point in time in the simulation is the contact between the infectious agents and those that are shielded, that is, those that are located at cell  $S^S$ . If the  $Q_t^C$  parameter requires the agent to isolate after having developed a clinical infection, then the agents located at the  $I^C$  cell should be excluded from the contact list too (lines 1-2). The algorithm proceeds with the infection spreading phase as outlined above (lines 3-19).

### 3. Results and discussion

To demonstrate the effectiveness of the CP-ABM methodology, we applied it to Northern Ireland, (one of the four countries that make up the United Kingdom). The dataset used consists of 251 days of observations of confirmed COVID-19 cases that were collected in the period between the 9<sup>th</sup> March and 15<sup>th</sup> November 2020 in Northern Ireland (NI). In this period the devolved NI government introduced and/or lifted many interventions with the goal of preventing and controlling the spread of disease. We used the CP component described in Section 2.1 to identify which of the decisions affecting social-mixing patterns that were the major drivers behind the changes in the COVID-19 infection rates. These key NPIs are used in the ABM component of the model (Section 2.2) to inform how infectious agents should adjust their contact behaviour as the simulation progresses. Table 2 contains a timeline of the occurrences of these interventions and at which tick the agents should change their mean daily contact behaviour in the simulation. In the simulation time period the two intervals in ticks (19, 144] and (221, 251] correspond to two national lockdowns that were introduced in March and October respectively, whereas the period between ticks (206, 221] correspond to a regional lockdown in the Derry and Strabane local government district (LGD) in October 2020. In NI the mandatory wearing of masks commenced on the 10<sup>th</sup> August 2020 and remained in place until the end of the observational study. The wearing of masks is incorporated into the model through its impact on the overall relative infectiousness, not mean daily contacts, and is captured by the behavioural parameter. An overview of all the behavioural and mean daily contact parameters that were used in the simulations is provided in Table 3. This requires fitting a total of 22 parameters, the process of which is described in Section 3.1. If we were interested in modelling a shorter time window, this would require fitting fewer  $m_{(\tau_i, \tau_j]}$  and  $b_{(\tau_i, \tau_j]}$  parameters, which would result in fitting a fewer number of parameters overall. This is entirely driven by the number of key NPIs identified by the CP component during the modelled time period. The timeline of these events plotted against the daily numbers of confirmed COVID-19 cases is illustrated in Fig. 4.

**Algorithm 2**

The pseudo-code of the Agent Based Modelling (ABM) component of the CP-ABM model.

INFECTIONSPREAD( $\tau$ )**Input:**


---

```

 $Q_\tau^C; \tau_b^I; b_\tau; m_\tau$ 
1.  $\mathcal{C}_\tau^I \leftarrow \{I^S, I^P\}, \text{ if } Q_\tau^C = \text{TRUE}$  Initialisation
    $\mathcal{C}_\tau^I \leftarrow \{I^S, I^P, I^C\}, \text{ if } Q_\tau^C = \text{FALSE}$ 
2.  $\mathcal{C}_\tau^I \leftarrow \{C \setminus \{S^S, I^C\}\}, \text{ if } Q_\tau^C = \text{TRUE}$ 
    $\mathcal{C}_\tau^I \leftarrow \{C \setminus \{S^S\}\}, \text{ if } Q_\tau^C = \text{FALSE}$ 
3. for each  $\zeta^I \in \mathcal{C}_\tau^I$  do Infection spreading
4.    $\mathcal{A}_\tau^I \leftarrow$  set of all agents located at the infectious cell  $\zeta^I$  at tick  $\tau$ 
5.   if  $\mathcal{A}_\tau^I \neq \emptyset$  then
6.     for each  $a^I \in \mathcal{A}_\tau^I$  do
7.        $N_\tau^{a^I} \leftarrow \text{Poisson}(m_\tau) \triangleright$  number of contacts agent  $a^I$  has at tick  $\tau$ 
8.        $\mathcal{A}_\tau^{a^I} \leftarrow$  set of  $N_\tau^{a^I}$  randomly selected agents from  $\mathcal{C}_\tau^I$  that are contacted by the agent  $a^I$ 
9.       for each  $a^{S'} \in \mathcal{A}_\tau^{a^I}$  do
10.        if  $\zeta^{S'} = S''$  then
11.           $P_a^{S \rightarrow E} \leftarrow \tau_b^I \cdot (1 - b_\tau) \cdot s_a^S$ 
12.          if  $\mathcal{U}(0, 1) < P_a^{S \rightarrow E}$  then
13.            Move agent  $a$  from cell "s" to cell "E" and set  $\delta_a^E$  to  $\Gamma(\mu = 3.0, k = 4)$  ticks
14.          end if
15.        end if
16.      end for
17.    end for
18.  end if
19. end for

```

---

**Table 2**

Timeline of the key NPIs extracted by the CP component that most impacts on social mixing.

| Date        | Tick ( $\tau$ ) | Non-Pharmaceutical Intervention  |
|-------------|-----------------|--|
| 28 Mar 2020 | 19              | From the following day the first full national lockdown was imposed restricting people by not allowing them to leave home without reasonable cause and forced businesses and schools to close. The restrictions also introduced the shielding programme for the elderly and vulnerable groups. |
| 18 May 2020 | 70              | Lockdown measures eased to allow a group of six to meet outdoors from the following day.   |
| 22 Jun 2020 | 105             | Further lockdown measures eased to allow up to six people to meet indoors from the following day.  |
| 02 Jul 2020 | 115             | Easing of the lockdown measures continues. From the following day all businesses in the hospitality sector (e.g. restaurants, hotels, pubs and cafes) are allowed to re-open.  |
| 31 Jul 2020 | 144             | From the following day the entire shielding programme is suspended. This day can be considered as the end of the first national lockdown.  |
| 23 Aug 2020 | 167             | Schools are reopened from the following day with pupils in grades seven, twelve and fourteen returning first.  |
| 13 Sep 2020 | 188             | From the following day, freshers week in NI universities started.  |
| 01 Oct 2020 | 206             | Due to increased number of confirmed COVID-19 cases, a localised circuit breaker lockdown was imposed on Derry and Strabane local government district from the following day limiting people's movement and reducing the operation of hospitality and leisure sectors.                         |
| 16 Oct 2020 | 221             | The circuit breaker lockdown was expanded to the entire nation including the closure of hospitality, close contact services and schools from the following day.  |
| 01 Nov 2020 | 237             | From the following day, the schools were permitted to reopen.  |

In order to model the COVID-19 infection dynamics as realistically as possible, the number of agents and their age characteristics used in the simulations conducted by the ABM component, have been set to match exactly the publicly available demographic projections for the NI population for 2019 [42], provided by the Northern Ireland Statistics and Research Agency (NISRA). More specifically, the simulation involved  $N^A = 1,893,667$  agents, of which 224,851 agents represented the elderly population aged 70+ years. At the beginning of the simulation, during the initialisation step ( $\tau = 0$ ), all non-elderly agents were randomly distributed across the cells as specified by the  $\eta_0^S$  parameter, where  $\zeta \in \mathcal{C}$  (Table 3). The parameter  $\eta_0^C$  was set to correspond to the number of confirmed clinical cases that were reported on the 9<sup>th</sup> March, and given that we take this date as a starting point in our simulation, we assumed that there is no recovered cases at that point of time and therefore no agents initialised at the cell with  $R$  status.

The shielding programme in NI was active between 29<sup>th</sup> March – 31<sup>st</sup> July (inclusive) and required the isolation of all vulnerable population groups such as people with comorbidities, terminal illness or those aged 70 and older. Given we do not take into account factors other than age for modelling the spread of disease,

to mimic real world behaviour in all of our simulations, the agents that represent the remaining susceptible elderly population enter the shielded state and move from cell with status  $S$  to cell with status  $S^S$  at tick  $\tau = 20$  and stay there until tick  $\tau = 144$  (inclusive). It is important to highlight that while shielded, the elderly agents are unable to be contacted by the infectious agents and thus unable to become infected. Upon the end of the shielding period, elderly agents become susceptible again and return back to the cell with status  $S$ .

To reflect the high level of adherence of the NI population to NPIs, we assumed that as soon as the agents develop clinical infection they go into isolation where they cannot be involved in any kind of interaction with others (i.e. cannot pass infection to others nor be contacted by other infectious agents), and hence it was specified that  $Q_\tau^C = \text{TRUE}$  for all  $\tau \in \{1, \dots, 251\}$ . This implies that in our simulations we assume that the infection is transmitted only via agents who developed either preclinical or subclinical infection. The true number of these agents at the beginning of simulation is unknown so we used calibration to define how many agents needs to be initialised at the cells with status  $I^P$  and  $I^S$ . All remaining agents (including all of those representing the elderly population) were initialised at the cell with status  $S$ .



**Table 3**

Overview of all input parameters and their values for the ABM component.

| Parameter               | Value   | Description  |
|-------------------------|---|--|
| $\tau_b^{\zeta^l}$      | $= \begin{cases} 1, & \text{if } \zeta^l = I^p \\ 0.39, & \text{if } \zeta^l = I^s \end{cases}$ | Baseline relative infectiousness of agents with developed subclinical, preclinical or clinical infection status <b>[calibrated]</b> .  |
| $p^{E \rightarrow I^s}$ | 0.38  | Probability of exposed agents developing a subclinical infection at the end of the incubation period. Otherwise the agent will develop a preclinical infection <b>[calibrated]</b> . |
| $m_{(0,19]}$            | 7.6   | Mean daily contacts of the agents on the segment $(\tau_i, \tau_j]$ <b>[calibrated]</b> .  |
| $m_{(19,70]}$           | 5.4   |  |
| $m_{(70,105]}$          | 5.5   |  |
| $m_{(105,115]}$         | 6.3   |  |
| $m_{(115,144]}$         | 6.4   |  |
| $m_{(144,167]}$         | 6.7   |  |
| $m_{(167,188]}$         | 6.8   |  |
| $m_{(188,206]}$         | 7.4   |  |
| $m_{(206,221]}$         | 7.1   |  |
| $m_{(221,237]}$         | 5.6   |  |
| $m_{(237,251]}$         | 5.7   |  |
| $b_{(0,19]}$            | 0   |  |
| $b_{(19,144]}$          | 0.18  |  |
| $b_{(144,153]}$         | 0   |  |
| $b_{(153,206]}$         | 0.4   |  |
| $b_{(206,221]}$         | 0.41  |  |
| $b_{(221,251]}$         | 0.48  |  |
| $\eta_0^S$              | 1,893,604   | Associated behavioural effect parameter of the agents on the segment $(\tau_i, \tau_j]$ <b>[calibrated]</b> .  |
| $\eta_0^{S^s}$          | 0   | Number of agents at cell S calculated on initialisation as $1,893,667 - (\eta_0^{S^s} + \eta_0^E + \eta_0^{I^s} + \eta_0^{I^p} + \eta_0^R)$ .  |
| $\eta_0^E$              | 29  | Number of agents at cell $S^s$ on initialisation <b>[observed]</b> .   |
| $\eta_0^{I^s}$          | 3   | Number of agents at cell E on initialisation <b>[calibrated]</b> .   |
| $\eta_0^{I^p}$          | 15  | Number of agents at cell $I^s$ on initialisation <b>[calibrated]</b> .   |
| $\eta_0^R$              | 16  | Number of agents at cell $I^p$ on initialisation <b>[calibrated]</b> .   |
| $\eta_0^R$              | 0   | Number of agents at cell $I^c$ on initialisation <b>[observed]</b> .   |
| $\eta_0^R$              | 0   | Number of agents at cell R on initialisation <b>[assumed]</b> .  |

### 3.1. Calibration and validation

For the purpose of the ABM component calibration we utilised the standard generational genetic algorithm (GA) [43] with population size 50, 15% mutation rate and 85% crossover rate, using tournament selection with tournament size 2. Each individual in the GA population is called a chromosome, and consists of a sequence of 20 value-encoded genes that represent values of input parameters for the ABM component. To address the stochastic nature of the ABM component and to identify the emerging infectious dynamic pattern that it produces, the simulation runs are repeated for 50 times (each time with a different random number generator seed) with the input parameter settings matching those of the corresponding chromosome. This enables the average fitness score to be calculated as follows:

$$\overline{MSE} = \frac{1}{50} \sum_{i=1}^{50} MSE(\hat{\mathcal{D}}_i, \mathcal{D}) \quad (2)$$

where  $MSE(\hat{\mathcal{D}}_i, \mathcal{D}) = \frac{1}{251} \sum_{j=1}^{251} (\hat{d}_{i,j} - d)^2$  is the mean square error of the difference between the estimated values of daily confirmed cases produced by the  $i^{th}$  simulation and the actual observed data. This fitness function describes how closely the ABM component fits the data, with lower values indicating a better fit, and is calculated for each chromosome in the population separately. The GA iterates from one population of chromosomes to another minimising the fitness function by giving preference to the chromosomes with lower fitness values which are used for reproduction during the tournament selection process, at the end of each generation.

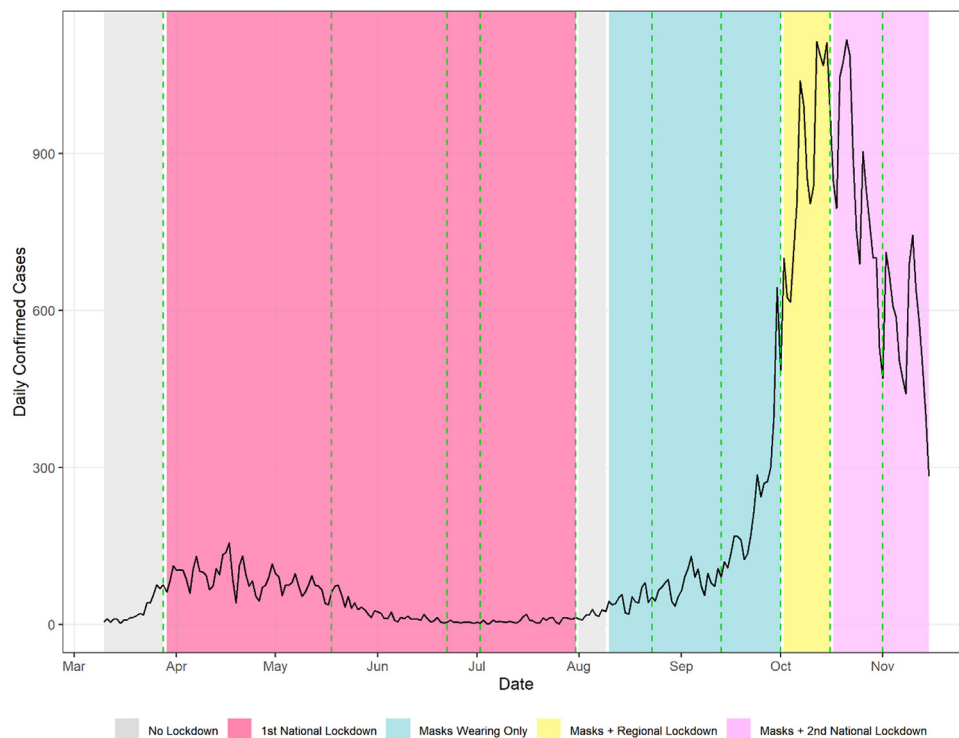
To identify the optimal values of the input parameters of the ABM component for the NI model, the GA was running continuously for almost 18 days and stopped when there was no further improvement in the fitness score. The fitness score of the best fit ABM component was  $\overline{MSE} = 10899.72$ . An overview of the input parameters, and their values that correspond to the best fitted model is provided in Table 3. All estimates produced by the ABM

component are obtained by running the simulations 50 times using the same random number generator seeds as the ones used by the GA with the best solution (a choice made for the reproducibility reasons), and averaging the results obtained.

In order to ensure that the ABM component is producing realistic patterns, deeming it fit for purpose, the patterns in its output had to agree with the patterns in the observed dataset, which is validated against several different criteria. Firstly, all GA estimated input parameters relating to the mean daily contacts  $m_{(\tau_i, \tau_j]}$  are within 95% confidence interval (CI) of the values reported by the independently conducted national COMIX study on social mixing patterns covering the same time period. Likewise, the ABM component is a good fit visually between the estimated cumulative and daily confirmed COVID-19 cases and the actual observations (Fig. 5a)). This is supported by the  $R^2$  and Pearson's  $R$  metrics that were in both instances higher than 0.93 and 0.96, respectively. Furthermore, the Kolmogorov-Smirnov test that was used to compare the distributions of the estimated and actual observations of weekly numbers of confirmed COVID-19 cases (Fig. 5b)) was non-significant ( $D = 0.28$ ,  $p = 0.12$ ). And finally, during the observed time period 47,162 cumulative confirmed COVID-19 cases were reported in total in NI, with the ABM component estimations being 47,164 with 95% CI = [44, 539; 49, 789] indicating that there is no statistically significant difference between the estimated and actual observations (Fig. 5 c)).

### 3.2. Interpretation of fitted parameters

Following the validation of the model, we can focus our attention on the other ABM input parameters that have been estimated by the GA algorithm. We have estimated that the overall proportion ( $p^{E \rightarrow I^s}$ ) of people who contract COVID-19 and develop a sub-clinical infection is 38%, which falls within the 95% prediction interval of the proportions reported in the study by Buitrago-Garcia et al. [5]. There are no clear guidelines on the base relative infectiousness of people with developed subclinical infection, however



**Fig. 4.** The number of new daily confirmed COVID-19 cases plotted alongside the timeline of NPIs identified by the CP component. The green dashed vertical lines indicate at which points the NPIs were made in this observation period (as outlined in Table 2) and the shaded backgrounds represent the duration of key NPIs such as the national and regional lockdowns, the wearing of facemasks and a period of no lockdown restrictions at all.

given that the model has proven to validate well against several different criteria and that the estimated  $P^{E \rightarrow I^S}$  is in line with the findings reported in the literature, we estimate that their baseline relative infectiousness  $\tau_b^S$  is 39%.

The main effects attributed to lockdowns is the change in social mixing patterns captured by the mean daily contacts parameter. In addition lockdowns can also lead to a change in other behavioural aspects of the population which has been captured by the behavioural effects parameter. We assume that these additional behavioural changes are a direct result of other NPIs that were not directly impacting social mixing patterns but did have an impact on keeping the disease spread under control. These measures include, but are not limited to the introduction of hand sanitising and 2 metre distancing rules in public places, special visors for employees, increased marketing campaigns and signage promoting hand hygiene.

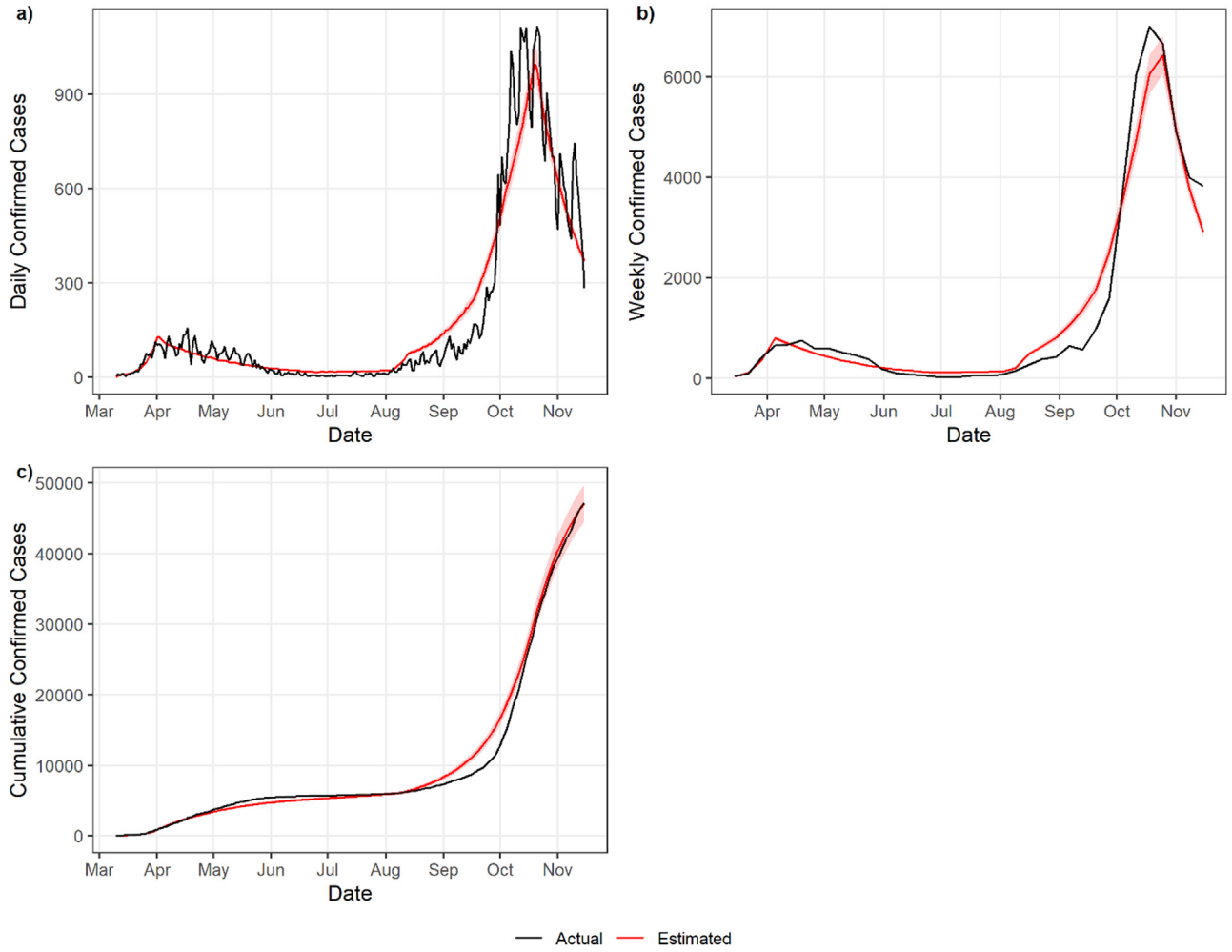
The introduction of the national lockdown at the beginning of the pandemic in April 2020 led to a sharp decrease in the mean daily contacts from 7.6 to 5.4 and likewise for the second national lockdown in the middle of October, reduced the mean daily contacts from 7.1 to 5.6. Aside from the lockdowns, the most striking change in the mean daily contacts occurs on 22<sup>nd</sup> June 2020 when there was a sharp increase from 5.5 to 6.3. According to the NPIs at the time, this corresponds directly to the relaxation that permitted up to 6 people from different households to meet indoors. The other most striking change is on the 13<sup>th</sup> September 2020 when the mean daily contacts increased from 6.8 to 7.4, corresponding to the return of students to Universities.

In addition to the reduction in the mean daily contacts, we should draw our attention to the behavioural effects parameters. During the first lockdown, the captured behavioural effects were attributed to the national lockdown only and reported as  $b_{(19,144)} = 0.18$ . After the first lockdown, the wearing of masks was introduced and made compulsory which is captured by the be-

havioural effects parameter. This wearing of masks continued to be compulsory for the duration of the study which means any further behavioural effects attributed to other lockdowns are in fact a mixed effect of wearing masks and lockdown. For a limited period of time (between the 10<sup>th</sup> August and the 1<sup>st</sup> October) there was the isolated effect attributed to only wearing masks which was captured by the parameter ( $b_{(153,206)} = 0.40$ ). With this in mind, the resulting behavioural effects parameter for the regional lockdown  $b_{(206,221)} = 0.41$  implies an additional 0.01 effect for the lockdown on top of the 0.40 effect for the masks and the previously reported reduction in mean daily contacts. Likewise, we can follow the same analogy for the second national lockdown where the behavioural effect of  $b_{(221,251)} = 0.48$  represents an additional 0.08 effect for the lockdown on top of the 0.40 effect for the masks and the previously reported reduction in the mean daily contacts.

### 3.3. Evaluation of NPI effectiveness

In the period between the 10<sup>th</sup> August (tick 154) and 1<sup>st</sup> October (tick 206) only the wearing of masks was compulsory (blue coloured area in Fig. 4) therefore, the number of cumulative confirmed COVID-19 cases in this time period can be used to quantify the effectiveness of wearing masks alone (when there is no mixed effect). The quantification of its effectiveness was performed by investigating what would be the estimated numbers of cumulative confirmed cases if the mask wearing rules were never introduced in the first place and comparing the estimated results with the actual data for the period between ticks 154 and 205 (inclusive). This can be achieved by setting the behavioural parameter corresponding to the time interval  $b_{(153,206)}$  to 0, and running a series of 50 ABM component simulations for the 206 ticks, using the same random number generator seeds as the ones used by the GA with the best solution (a choice made for the reproducibility reasons). The estimated numbers of cumulative confirmed cases is averaged



**Fig. 5.** The estimated number of new COVID-19 cases against time compared with the actual observed number of (a) daily, (b) weekly, and (c) cumulative confirmed COVID-19 cases.

across for all simulations and the percentage increase on the  $i^{\text{th}}$  day ( $\Delta_i$ ) calculated relative to the actual data using:

$$\Delta_i = \frac{\bar{\lambda}_i - \lambda_i}{\lambda_i} \times 100 \quad (3)$$

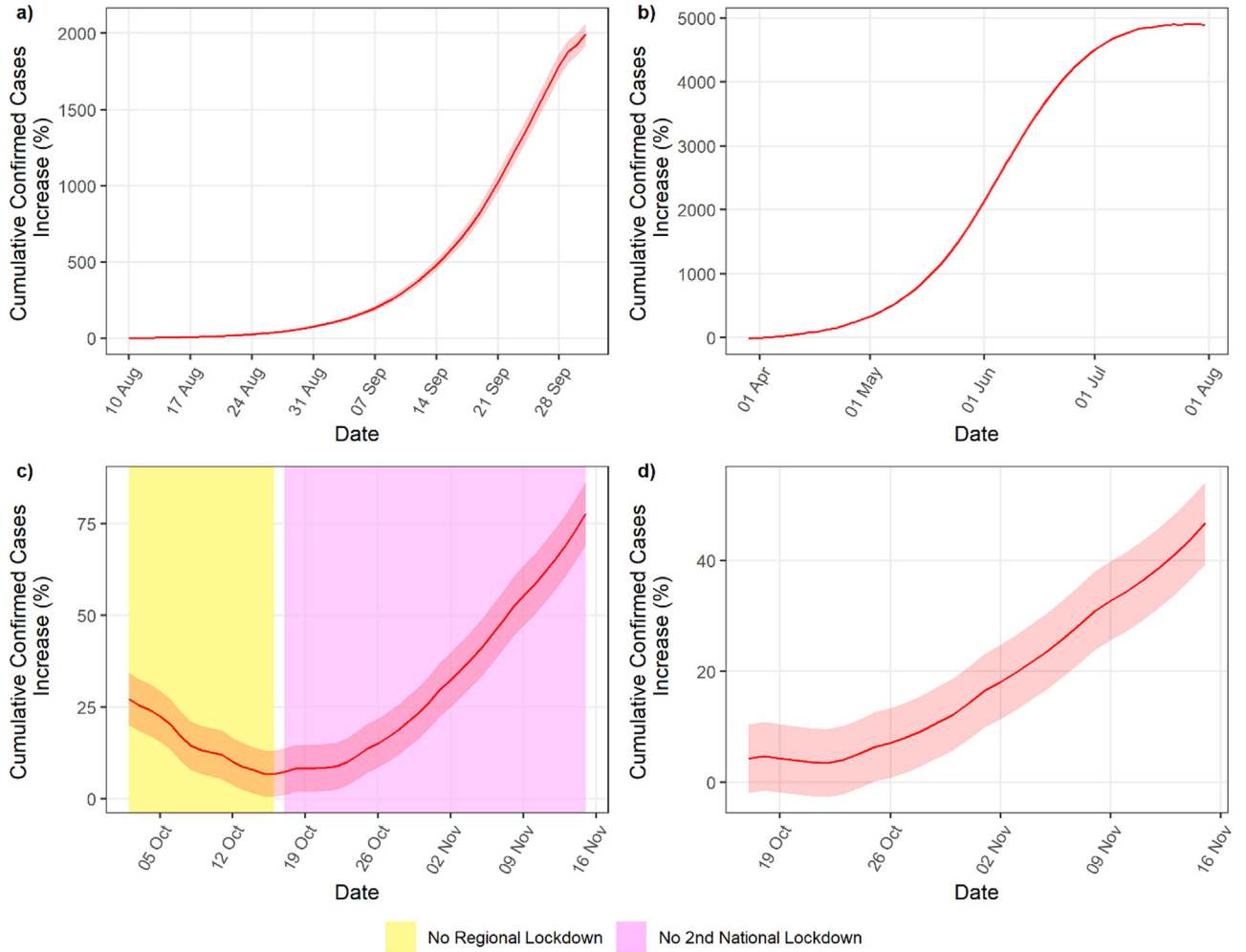
where  $\lambda_i$  and  $\bar{\lambda}_i$  are the actual and projected numbers of cumulative confirmed cases respectively.

These results have been further validated by comparing the CP-ABM estimations with the ones found in the literature. In Germany, Mitze et al. [44] suggested that 20 days after becoming mandatory face masks reduced the number of new COVID-19 infections by 45%. This is almost identical results to what we have reported from the CP-ABM model where we estimate approximately a 46% increase in the number of reported cases on the 28<sup>th</sup> August, the 19<sup>th</sup> day after mask wearing became mandatory. We extend Mitze et al.'s findings by estimating the impact beyond 20 days noting that the number of reported cases would double to what was reported after only 24 days, and more than triple after only 30 days into the estimation time period. After that, with each additional day that passes, the estimated daily number of confirmed cases rapidly grows towards a near complete spread with the projected numbers being almost 21 times higher on the 1<sup>st</sup> October compared to the actual number that was reported in NI at the time when mask wearing did take place. This translates to 267,915 (95% CI = [258,861; 276,969]) estimated confirmed cases on the 1<sup>st</sup> October if masks were never introduced compared to 12,792

confirmed cases of what was actually observed. Daily cumulative growth projections of the number of confirmed COVID-19 cases in the scenario when masks were never introduced is illustrated in Fig. 6 a).

A similar approach was taken for quantifying the effectiveness of the lockdowns. Given the assumption that the NPIs being made during lockdowns lead to changes in social-mixing patterns and other aspects of people's behaviour, the effectiveness of the lockdowns was estimated by altering both the mean daily contacts and behavioural effect parameters. Experiments are then simulated to consider what would have happened if lockdowns were never introduced by setting the mean daily contacts and behavioural parameters during the lockdown to be equal to their respective values prior to lockdown and kept constant for the entire evaluation period.

We conducted three experiments in total. In the **first experiment** we explore the impact of the first national lockdown by projecting the number of cumulative confirmed cases in the scenario in which the first national lockdown was never introduced and hence perform the simulations when the mean daily contacts and behavioural effects parameters are set to match those prior the lockdown, i.e.  $m_{(19,144)} = m_{(0,19)} = 7.6$  and  $b_{(19,144)} = b_{(0,19)} = 0$ , respectively. In the **second experiment** we explore the impact of the regional lockdown by projecting the number of cumulative confirmed cases in the scenario in which the regional lockdown and



**Fig. 6.** The estimated percentage increase in cumulative confirmed COVID-19 cases (Eq. (3)) in Northern Ireland if: (a) the masks wearing rules were never introduced (b) there was no first national lockdown, (c) there was no regional lockdown in Derry and Strabane local government district nor second national lockdown, and (d) there was no second national lockdown.

second national lockdown did not occur. This required setting the mean daily contacts to be the same as the ones that preceded the regional lockdown  $m_{(206,221)} = m_{(221, 251)} = m_{(188,206)} = 7.4$ . As the behavioural effects of these two lockdowns are mixed effects, we nullified the part attributed to the lockdown and left the remainder that equated to only wearing masks assuming that  $b_{(206,221)} = b_{(221, 251)} = b_{(188,206)} = 0.4$ . Finally, in the **third experiment** we explore the impact of the second national lockdown by projecting the number of cumulative confirmed cases if the second national lockdown was never introduced, with only the regional lockdown taking place. Therefore, the mean daily contacts and behavioural effects parameters were set to match those of the regional lockdown, i.e.  $m_{(221,251)} = m_{(206,221)} = 7.1$  and  $b_{(221,251)} = b_{(206,221)} = 0.41$ .

For all three experiments the simulations were run in the same manner as when we estimated the effectiveness of the masks. The simulations for the first experiment was set up to run until tick 144, corresponding to the 31<sup>st</sup> July (the end of the first lockdown period), and for the other two experiments until tick 251, which is the 15<sup>th</sup> November (the end of observed time period). The corresponding projected percentage increase in the number of cumulative confirmed cases relative to the observed data was calculated using Eq. (3). The projections for the first experiment focus on the time interval concerning the first lockdown, illustrated in Fig. 6 b). Similarly, the projections for the second experiment covers the time interval that overlaps with both the regional and sec-

ond national lockdowns, illustrated in Fig. 6c). The projections for the third experiment are based on the time interval that concerns the second national lockdown, as shown in the Fig. 6d).

In the first experiment, if there was no national lockdown (Fig. 6b)), we would have seen nearly 50 times more confirmed cases on the 31<sup>st</sup> July than what was actually reported in NI on that date. To be specific, the estimated number of confirmed cases would have been 295,252 (95% CI = [294,838; 295,666]) compared to the 5,917 confirmed cases reported at that time in the actual dataset. For the second and third experiments, the projections are not as stark due to the ongoing behavioural effect associated with wearing masks that was not present in the first lockdown. In the case that there was neither a regional nor second national lockdown, the estimated number of cumulative confirmed cases would have been almost 78% greater in comparison to the numbers that have been reported on the 15<sup>th</sup> November in NI (Fig. 6 c)). This translates to 83,818 cases (95% CI = [79,750; 87,886]) estimated in comparison to 47,162 confirmed cases on the 15<sup>th</sup> November in NI. From Fig. 6 d) we can observe that having only a regional lockdown in the Derry and Strabane local government district was not a strong enough intervention on its own to prevent the entire spread of disease throughout the country, however it did have an impact on slowing the rise in cases as the projected number of cumulative confirmed cases is almost 47% as opposed to 78% greater in comparison to having no regional nor second lockdown (Fig. 6



c)). The figures also demonstrate that without the second national lockdown being introduced, we would have expected an increase of 47% in the number of confirmed cases on the 15<sup>th</sup> November.

### 3.4. Implementation details

The NI COVID-19 infection dynamics model that was built using the CP-ABM methodology, that we introduce in this paper, is implemented using a combination of multiple technologies. The entire CP-component algorithm and statistical data analytics of all the results is implemented using the R v4.0.3 programming language. The ABM component of the model, is implemented using the NetLogo v6.1.1 programming language. The calibration of input parameters was performed using the BehaviorSearch tool [45] which interfaces with NetLogo to automate the exploration of the ABM parameter spaces using the GAs [46]. The automatised pipeline that connects all of these tools together was implemented using the Microsoft PowerShell v5.1 programming language.

The GA calibration and NPI effectiveness quantification experiments presented in this paper were executed on a HP Z8 desktop workstation with 44 CPU (Intel Xeon) cores and 1.5TB of DDR4 RAM. To reduce the running time, all experiments were parallelised and executed one per CPU core. The average running time of the ABM component simulations depends on the values of its input parameters. The calibrated ABM component simulation, with input parameters as in Table 3, can run on any common grade hardware that has the NetLogo programming language installed. The average running time per simulation was about 123 seconds, as tested on the Dell Precision 5750 laptop with Intel i7-10875H CPU with 16 CPU cores and 64 GB of DDR-4 RAM. However using the parallelised approach to run 50 simulations with the same input parameters across all 16 cores takes approximately 45 minutes to completion. These performances significantly improve what was reported in the literature [26,27] and previously outlined in Section 1.1.2.

## 4. Conclusion

This paper introduces the CP-ABM, a new state of the art methodology that is capable of accurately mimicking the disease dynamics for COVID-19. The application of the CP-ABM to the Northern Ireland population is included as a demonstration of its suitability and effectiveness for this purpose. The model covers the time period of 251 days and to the best of our knowledge, this study is the first of its kind that was able to successfully capture COVID-19 infection dynamics during such a long period of time and successfully model both infection waves. Genetic algorithms were used for calibration purposes to ensure that the exact infection dynamics patterns can be replicated by the CP-ABM. This was validated through the existing literature. We also capture the role played by the subclinical people in the infection dynamics workflow.

The CP-ABM approach uniquely incorporates change point detection into agent based models in order to identify key events which leads to changes in contact behaviour between people which is the driving force behind the COVID-19 infection rates [43]. The ABM component, unlike the top down approaches of the equation based modelling techniques, allows the simulation of each individual's behaviour in the population. The resulting key events from the CP component are used in the ABM component to appropriately direct the adaptive behaviour of its agents to produce realistic results. The demand for high computational power has limited the application of ABMs on a population level for COVID-19. However, we overcome this issue by employing an efficient infection centric modelling approach which aims to capture the interactions of only those agents who are able to transmit the disease to

others during the simulation. The proposed methodology enables researchers to develop models that can be run and calibrated on consumer grade hardware as it is programming language agnostic and does not necessarily require supercomputers.

The new CP-ABM methodology is also able to quantify the effects of non-pharmaceutical interventions on a national level which we demonstrated in use cases that estimate the effectiveness of masks and regional and national lockdowns. This has been achieved through the simulations of different scenarios which consider what would have happened if those NPIs were never introduced, and compare the estimated projections of the confirmed daily cases with the real observations. We identify that the NPIs in Northern Ireland considered in our experiments were necessary and highly effective in preventing an even more extreme outbreak of the country's pandemic.

Although not explored in this paper, the model can be expanded to have the functionality to capture the number of agents on each cell and their interactions at each tick during simulations. This can be utilised in future work to consider the interaction between agents in different settings and to evaluate scenarios for considering the impact of vaccines, new COVID-19 variants and detailed analysis of clusters dynamically changing over time. We demonstrated the CP-ABM capabilities through the implementation of the model on the NI population but the model is not unique to NI and therefore it can be applied to other countries and geographical areas (both large and small). Also, the methodology is not COVID-19 specific and hence may be used as a basis for creating epidemiological models to capture infection dynamics properties and the assessment of the NPI effects of other diseases both in human and animal health.

Our research is implemented using the Netlogo language which is a higher level language for rapid development of agent based models. The speed could be significantly improved if the algorithm was implemented in one of the lower level languages, such as C, C++ or Rust. Also in this paper, we use standard genetic algorithms for the purpose of the ABM component's hyperparameters calibration and in the future we wish to explore other optimisation strategies to investigate whether they would lead to obtaining faster results.

## Funding

This research has been funded by Queen's University Belfast.

## 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.

## Acknowledgement

The authors wish to acknowledge the Northern Ireland COVID-19 Modelling Group for their interaction in the model building process and their valuable feedback. We also wish to acknowledge Advanced Analytics Labs Ltd for the provision of the computing resources for model building, calibration and running of the experiments described in this paper.

## References

- [1] WHO, *Transmission of SARS-CoV-2: implications for infection prevention precautions*, (n.d.). <https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions> (accessed December 19, 2020).
- [2] S.A. Lauer, K.H. Grantz, Q. Bi, F.K. Jones, Q. Zheng, H.R. Meredith, A.S. Azman, N.G. Reich, J. Lessler, The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application, *Ann. Intern. Med.* 172 (2020) 577–582, doi:10.7326/M20-0504.

- [3] L.C. Tindale, J.E. Stockdale, M. Coombe, E.S. Garlock, W.Y.V. Lau, M. Saraswat, L. Zhang, D. Chen, J. Wallinga, C. Colijn, Evidence for transmission of covid-19 prior to symptom onset, *Elife* 9 (2020) 1–34, doi:[10.7554/ELIFE.57149](https://doi.org/10.7554/ELIFE.57149).
- [4] K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020, *Eurosurveillance* 25 (2020) 2000180, doi:[10.2807/1560-7917.ES.2020.25.10.2000180](https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180).
- [5] D. Buitrago-Garcia, D. Egli-Gany, M.J. Counotte, S. Hossmann, H. Imeri, A.M. Ipekci, G. Salanti, N. Low, Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: a living systematic review and meta-analysis, *PLoS Med.* 17 (2020) e1003346, doi:[10.1371/journal.pmed.1003346](https://doi.org/10.1371/journal.pmed.1003346).
- [6] K. Prem, Y. Liu, T.W. Russell, A.J. Kucharski, R.M. Eggo, N. Davies, S. Flasche, S. Clifford, C.A.B. Pearson, J.D. Munday, S. Abbott, H. Gibbs, A. Rosello, B.J. Quilty, T. Jombart, F. Sun, C. Diamond, A. Gimma, K. van Zandvoort, S. Funk, C.I. Jarvis, W.J. Edmunds, N.I. Bosse, J. Hellewell, M. Jit, P. Klepac, The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study, *Lancet Public Heal* 5 (2020) e261–e270, doi:[10.1016/S2468-2667\(20\)30073-6](https://doi.org/10.1016/S2468-2667(20)30073-6).
- [7] M.H.D.M. Ribeiro, R.G. da Silva, V.C. Mariani, L. dos S. Coelho, Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil, *Chaos Solitons Fract.* 135 (2020) 109853, doi:[10.1016/j.chaos.2020.109853](https://doi.org/10.1016/j.chaos.2020.109853).
- [8] P.C.L. Silva, P.V.C. Batista, H.S. Lima, M.A. Alves, F.G. Guimarães, R.C.P. Silva, COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions, *Chaos Solitons Fract.* 139 (2020) 110088, doi:[10.1016/j.chaos.2020.110088](https://doi.org/10.1016/j.chaos.2020.110088).
- [9] A. Bouchnita, A. Jebrane, A hybrid multi-scale model of COVID-19 transmission dynamics to assess the potential of non-pharmaceutical interventions, *Chaos Solitons Fract.* 138 (2020) 109941, doi:[10.1016/j.chaos.2020.109941](https://doi.org/10.1016/j.chaos.2020.109941).
- [10] E. Hunter, B. Mac Namee, J. Kelleher, A hybrid agent-based and equation based model for the spread of infectious diseases, *JASSS* 23 (2020) 1–25, doi:[10.18564/jasss.4421](https://doi.org/10.18564/jasss.4421).
- [11] J. Tolles, T. Luong, Modeling epidemics with compartmental models, *JAMA - J. Am. Med. Assoc.* 323 (2020) 2515–2516, doi:[10.1001/jama.2020.8420](https://doi.org/10.1001/jama.2020.8420).
- [12] W. Duan, Z. Fan, P. Zhang, G. Guo, X. Qiu, Mathematical and computational approaches to epidemic modeling: a comprehensive review, *Front. Comput. Sci.* 9 (2015) 806–826, doi:[10.1007/s11704-014-3369-2](https://doi.org/10.1007/s11704-014-3369-2).
- [13] M.Y. Li, J.R. Graef, L. Wang, J. Karsai, Global dynamics of a SEIR model with varying total population size, *Math. Biosci.* 160 (1999) 191–213, doi:[10.1016/S0025-5564\(99\)00030-9](https://doi.org/10.1016/S0025-5564(99)00030-9).
- [14] T. Kuniya, Prediction of the epidemic peak of coronavirus disease in Japan, 2020, *J. Clin. Med.* 9 (2020) 789, doi:[10.3390/jcm9030789](https://doi.org/10.3390/jcm9030789).
- [15] F. Ndaïrou, I. Area, J.J. Nieto, D.F.M. Torres, Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan, *Chaos Solitons Fract.* 135 (2020) 109846, doi:[10.1016/j.chaos.2020.109846](https://doi.org/10.1016/j.chaos.2020.109846).
- [16] N.G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, CMMID COVID-19 working group, R.M. Eggo, Age-dependent effects in the transmission and control of COVID-19 epidemics, *Nat. Med.* 26 (2020) 1205–1211, doi:[10.1038/s41591-020-0962-9](https://doi.org/10.1038/s41591-020-0962-9).
- [17] A.R. Tuite, D.N. Fisman, A.L. Greer, Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada, *CMAJ* 192 (2020) E497–E505, doi:[10.1503/cmaj.200476](https://doi.org/10.1503/cmaj.200476).
- [18] Z. Wang, Y. Xiao, Y. Li, J. Zhang, F. Lu, M. Hou, X. Liu, Automatically discriminating and localizing COVID-19 from community-acquired pneumonia on chest X-rays, *Pattern Recognit.* 110 (2021) 107613, doi:[10.1016/j.patcog.2020.107613](https://doi.org/10.1016/j.patcog.2020.107613).
- [19] M. Shorfuzzaman, M.S. Hossain, MetaCOVID: A Siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients, *Pattern Recognit.* 113 (2021) 107700, doi:[10.1016/j.patcog.2020.107700](https://doi.org/10.1016/j.patcog.2020.107700).
- [20] J. Li, G. Zhao, Y. Tao, P. Zhai, H. Chen, H. He, T. Cai, Multi-task contrastive learning for automatic CT and X-ray diagnosis of COVID-19, *Pattern Recognit.* 114 (2021) 107848, doi:[10.1016/j.patcog.2021.107848](https://doi.org/10.1016/j.patcog.2021.107848).
- [21] K.E. Arunkumar, D.V. Kalaga, C.M.S. Kumar, M. Kawaji, T.M. Brenza, Forecasting of COVID-19 using deep layer Recurrent Neural Networks (RNNs) with Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) cells, *Chaos Solitons Fract.* 146 (2021) 110861, doi:[10.1016/j.chaos.2021.110861](https://doi.org/10.1016/j.chaos.2021.110861).
- [22] C.C. John, V. Ponnusamy, S. Krishnan Chandrasekaran, N. Ra, A survey on mathematical, machine learning and deep learning models for COVID-19 transmission and diagnosis, *IEEE Rev. Biomed. Eng.* 15 (2022) 325–340, doi:[10.1109/RBME.2021.3069213](https://doi.org/10.1109/RBME.2021.3069213).
- [23] O. Ozmen, J.J. Nutaro, L.L. Pullum, A. Ramanathan, Special section on medical simulation analyzing the impact of modeling choices and assumptions in compartmental epidemiological models, *Model. Simul. Int.* 92 (2016) 459–471, doi:[10.1177/0037549716640877](https://doi.org/10.1177/0037549716640877).
- [24] F. Miksch, B. Jahn, K.J. Espinosa, J. Chhatwal, U. Siebert, N. Popper, Why should we apply ABM for decision analysis for infectious diseases?—An example for dengue interventions, *PLoS One* 14 (2019) e0221564, doi:[10.1371/journal.pone.0221564](https://doi.org/10.1371/journal.pone.0221564).
- [25] L. Perez, S. Dragicevic, An agent-based approach for modeling dynamics of contagious disease spread, *Int. J. Health Geogr.* 8 (2009) 50, doi:[10.1186/1476-072X-8-50](https://doi.org/10.1186/1476-072X-8-50).
- [26] S.L. Chang, N. Harding, C. Zachreson, O.M. Cliff, M. Prokopenko, Modelling transmission and control of the COVID-19 pandemic in Australia, *Nat. Commun.* 11 (2020) 1–13, doi:[10.1038/s41467-020-19393-6](https://doi.org/10.1038/s41467-020-19393-6).
- [27] R. Hinch, W.J.M. Probert, A. Nurtay, M. Kendall, C. Wymant, M. Hall, K. Lythgoe, A.B. Cruz, L. Zhao, A. Stewart, L. Ferretti, D. Montero, J. Warren, N. Mather, M. Abueg, N. Wu, A. Finkelstein, D.G. Bonsall, L. Abeler-Dörner, C. Fraser, OpenABM-Covid19 – An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing, *MedRxiv* (2020), doi:[10.1101/2020.09.16.20195925](https://doi.org/10.1101/2020.09.16.20195925).
- [28] N. Mahdizadeh Gharakhanlou, N. Hooshangi, Spatio-temporal simulation of the novel coronavirus (COVID-19) outbreak using the agent-based modeling approach (case study: Urmia, Iran), *Inform. Med. Unlocked.* 20 (2020) 100403, doi:[10.1016/j.imu.2020.100403](https://doi.org/10.1016/j.imu.2020.100403).
- [29] Y. Vyklyuk, M. Manylich, M. Škoda, M.M. Radovanović, M.D. Petrović, Modeling and analysis of different scenarios for the spread of COVID-19 by using the modified multi-agent systems – Evidence from the selected countries, *Results Phys.* 20 (2021) 103662, doi:[10.1016/j.rinp.2020.103662](https://doi.org/10.1016/j.rinp.2020.103662).
- [30] N. Hoertel, M. Blachier, C. Blanco, M. Olsson, M. Massetti, M.S. Rico, F. Limosin, H. Leleu, A stochastic agent-based model of the SARS-CoV-2 epidemic in France, *Nat. Med.* 26 (2020) 1417–1421, doi:[10.1038/s41591-020-1001-6](https://doi.org/10.1038/s41591-020-1001-6).
- [31] E. Cuevas, An agent-based model to evaluate the COVID-19 transmission risks in facilities, *Comput. Biol. Med.* 121 (2020) 103827, doi:[10.1016/j.combiomed.2020.103827](https://doi.org/10.1016/j.combiomed.2020.103827).
- [32] F. Araya, Modeling the spread of COVID-19 on construction workers: An agent-based approach, *Saf. Sci.* 133 (2021) 105022, doi:[10.1016/j.ssci.2020.105022](https://doi.org/10.1016/j.ssci.2020.105022).
- [33] C.I. Jarvis, K. Van Zandvoort, A. Gimma, K. Prem, M. Auzenberg, K. O'Reilly, G. Medley, J.C. Emery, R.M.G.J. Houben, N. Davies, E.S. Nightingale, S. Flasche, T. Jombart, J. Hellewell, S. Abbott, J.D. Munday, N.I. Bosse, S. Funk, F. Sun, A. Endo, A. Rosello, S.R. Procter, A.J. Kucharski, T.W. Russell, G. Knight, H. Gibbs, Q. Leclerc, B.J. Quilty, C. Diamond, Y. Liu, M. Jit, S. Clifford, C.A.B. Pearson, R.M. Eggo, A.K. Deol, P. Klepac, G.J. Rubin, W.J. Edmunds, Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK, *BMC Med.* 18 (2020) 124, doi:[10.1186/s12916-020-01597-8](https://doi.org/10.1186/s12916-020-01597-8).
- [34] I.A. Eckley, P. Fearnhead, R. Killick, Analysis of changepoint models, in: *Bayesian Time Ser. Model.*, Cambridge University Press, 2011, pp. 205–224, doi:[10.1017/CBO9780511984679.011](https://doi.org/10.1017/CBO9780511984679.011).
- [35] K. Haynes, I.A. Eckley, P. Fearnhead, Computationally efficient changepoint detection for a range of penalties, *J. Comput. Graph. Stat.* 26 (2017) 134–143, doi:[10.1080/10618600.2015.1116445](https://doi.org/10.1080/10618600.2015.1116445).
- [36] B. Blasius, Power-law distribution in the number of confirmed COVID-19 cases, *Chaos* 30 (2020) 093123, doi:[10.1063/5.0013031](https://doi.org/10.1063/5.0013031).
- [37] B.T. Fagan, M.I. Knight, N.J. MacKay, A.J. Wood, Change point analysis of historical battle deaths, *J. R. Stat. Soc. Ser. A (Statistics Soc.)* 183 (2020) 909–933, doi:[10.1111/rssa.12578](https://doi.org/10.1111/rssa.12578).
- [38] K. Haynes, P. Fearnhead, I.A. Eckley, A computationally efficient nonparametric approach for changepoint detection, *Stat. Comput.* 27 (2017) 1293–1305, doi:[10.1007/s11222-016-9687-5](https://doi.org/10.1007/s11222-016-9687-5).
- [39] N.R. Zhang, D.O. Siegmund, A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data, *Biometrics* 63 (2007) 22–32, doi:[10.1111/j.1541-0420.2006.00662.x](https://doi.org/10.1111/j.1541-0420.2006.00662.x).
- [40] J. Hilton, M.J. Keeling, Estimation of country-level basic reproductive ratios for novel Coronavirus (SARS-CoV-2/COVID-19) using synthetic contact matrices, *PLOS Comput. Biol.* 16 (2020) e1008031, doi:[10.1371/journal.pcbi.1008031](https://doi.org/10.1371/journal.pcbi.1008031).
- [41] S.F. Railsback, V. Grimm, *Agent-Based and Individual-Based Modeling: A Practical Introduction*, 2nd ed., Princeton University Press, Princeton, NJ, 2019.
- [42] Northern Ireland Statistics and Research Agency, *Home - NINIS: Northern Ireland Neighbourhood Information Service*, (n.d.). <https://www.ninis2.nisra.gov.uk/public/Home.aspx> (accessed March 24, 2021).
- [43] J.H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, MI, 1975 Ann Arbor.
- [44] T. Mitze, R. Kosfeld, J. Rode, K. Walde, Face masks considerably reduce COVID-19 cases in Germany, *Proc. Natl. Acad. Sci. U. S. A.* 117 (2020) 32293–32301, doi:[10.1073/pnas.2015954117](https://doi.org/10.1073/pnas.2015954117).
- [45] F. Stonedahl, U. Wilensky, BehaviorSearch [computer software], Cent. Connect. Learn. Comput. Based Model. Northwest. Univ. Evanston (March 25, 2021) n.d <http://www.behaviorsearch.org/> (accessed).
- [46] F.J. Stonedahl, Genetic algorithms for the exploration of parameter spaces in agent-based models, 2011.

**Dr Aleksandar Novakovic** is a Lecturer in Data Analytics at Queen's University Belfast and Adjunct Professor at Ontario Tech University with expertise in the fields of data analytics and artificial intelligence and their interdisciplinary applications in solving big data problems, especially the ones related to real-time healthcare analytics, and decision making under uncertainty.

**Professor Adele Marshall** is Professor of Statistics at Queen's University Belfast, Adjunct Professor at Ontario Tech University, and world leader in her field of research in statistical modelling and decision making under uncertainty primarily applied in healthcare. Since gaining her PhD in Statistics in 2001, Adele has co-authored over 200 publications of her research.