EXTENDED COMPARTMENTAL MODEL OF COVID-19: A CELL-DEVS DEFITINION

Aidan Fahlman
Systems and Computer Engineering
Carleton University
Ottawa, ON K1S 5B6, Canada
aidanfahlman@cmail.carleton.ca

Cristina Ruiz-Martin
Systems and Computer Engineering
Carleton University
Ottawa, ON K1S 5B6, Canada
cristinaruizmartin@sce.carleton.ca

Gabriel Wainer
Systems and Computer Engineering
Carleton University
Ottawa, ON K1S 5B6, Canada
gwainer@sce.carleton.ca

Peter Dobias
Defence Research and Development Canada
Centre for Operational Research and Analysis
Ottawa, ON, K1A0K2, Canada
peter.dobias@forces.gc.ca

Mark Rempel
Defence Research and Development Canada
Centre for Operational Research and Analysis
Ottawa, ON, K1A0K2, Canada
mark.rempel@forces.gc.ca

Abstract—Susceptible-Infected-Recovered (SIR) models have been used to study the spread of COVID-19. In previous works, the standard SIR model has been expanded to include new states as well as geographical level transmission dynamics. We present an extended model using the Cell-DEVS formalism that simulates the effect asymptomatic COVID-19 cases have on a population. The model is an easily adaptable framework that allows for rapid-prototyping and modifications. We exemplify how to build and easily change the model using public health units of Ontario as a case study. The results show the effect asymptomatic carriers have on overall case counts and exposures at the provincial level as well as at the city level.

Keywords—Simulation, Cell-DEVS, Asymptomatic, pandemic

I. INTRODUCTION

COVID-19 has been at the forefront of public health since late 2019. Despite the efforts of health agencies, in June 2021 many countries are coping with second and third waves of COVID-19 cases [1]. These new waves, often associated with variants, are typically reaching higher peaks than the first wave in 2020 [1], [2]. Some of the research into the causes of secondary waves suggests that they could be due to the asymptomatic carriers—that is those where the individual does not show symptoms and thus may be unaware that they are infected, do not take precautions [3], and thus may spread the disease to at risk individuals unknowingly. The proportion of asymptomatic carriers can be as high as 80% [3]–[7]. Short of regular testing of the entire population, which did not prove overly effective in at least some countries attempting it, one of the ways of estimating the prevalence of asymptomatic carriers is through modeling and simulation [8].

Health agencies and research organizations have long relied on modelling and simulations e.g., predictive models provide estimates of future trends; to assess potential effectiveness of various disease control methods such as lock downs, mandatory quarantines, social distancing [9]; etc. Of a particular usefulness can be so called geographical models which allow for geo-located simulations of various phenomena; their resolution typically range from continental models to city neighborhoods [10]. These disease control models use different inputs that characterize the disease being modelled, which can change depending on the evolution of a disease. Rapid prototyping allows for modelers to add new characteristics, as well as adapt the model to fit their evolving needs, with minimal changes. These changes should be efficient and accessible. This approach also enables one to incorporate additional critical parameters not known when the modelling commenced.

One of the most popular epidemiological models, introduced by Kermack and McKendrick in 1927 [11], classified the population in "compartments": Susceptible to the disease, Infective (i.e., can transmit the disease), and Recovered (SIR). Modifying the parameters of this type of model, public health officials can investigate how much a disease might continue to rise depending on the public health measures put in place. They can use these simulation results to plan public health measures before case counts reach an uncontrollable point. In these models, we can create a new compartment to incorporate asymptomatic infections and predict the impact they will have on symptomatic case counts.

SIR-type geographical models have been the forefront of disease growth and spread prediction. With the emergence of the SARS-Cov-2 virus, these models have become relevant, being used to track and monitor the potential spread of the disease. For instance, each component in city areas can use its own defined population and characteristics, allowing for accurate representation of a given geography. For example, [7] shows the negative effects of asymptomatic COVID-19 infections within homeless shelters. Our model could be used to geographically locate homeless shelters within specific neighborhoods and simulate the effect homeless shelters on surrounding neighborhoods. This idea can be adapted for any geographical level, from small city neighborhoods to state/provincial, or even larger simulations. Geographical modelling allows to simulate what regions are most impacted by a given infection. These insights can help create improved public health measures. Rapid prototyping allows for public

978-1-6654-3326-6/21/$31.00 ©2021 Crown

health officials to efficiently change the characteristics of a disease and re-run simulations showing different scenarios depending on the given characteristics. These changes allow for rapid prototyping a virus as it evolves over time.

The main contribution of this research is the definition of an extension of the traditional SIR model, specifically a geographical Susceptible-Exposed-Asymptomatic-Infective-Recovered-Deceased (SEAIRD) model whose aim is to incorporate advanced behavior for COVID-19. Our implementation allows users to run the model at a user-defined region level and visualize how COVID-19 might spread through a city, town, or country. The model is designed using the Cell-DEVS formalism [12] and implemented using the Cadmium simulator [13]. The model's adaptable framework allows for accessible rapid-prototyping and modifications. We implemented the asymptomatic cases using a basic infectious/asymptomatic ratio value that can be input by the user. The asymptomatic carriers expose more individuals than their non-asymptomatic counterparts. This is intended to represent the effect of not taking adequate precautions due to the absence of awareness. Users can also input a described group of neighborhoods and run the model through the neighborhoods, allowing for visualization of how a disease might spread through a city, town, or country. We use a case study where the model's neighborhoods are defined as the public health units of Ontario. Our results show how an asymptomatic set of carriers can lead to sharper increases in case counts resulting in a change to the total numbers of cases that a population would experience. Our model provides a framework to rapidly prototype disease spread in their neighborhood where asymptomatic infections can be considered and incorporated where necessary.

II. Background

A. Introduction to SIR-type Models

Following Ross and Hudson [14]–[16], Kermack and McKendrick [11] defined a model that classified a given population into three "compartments": Susceptible, Infectious, and Recovered (SIR). They defined how individuals within a population could move from one compartment to another over time. Kermack and McKendrick's work defined the framework and mathematics that SIR-type models continue to follow today. This standard SIR model has evolved over the years to incorporate more advanced disease spread rules and more compartments. The simplest of these evolutions is the SIRD model which incorporates a Deceased (D) compartment and includes death factors and fatality rates [17]. SEIRD models add an Exposed (E) state used as a transition from susceptible to infected [18]. Over time, these models became significantly more advanced, having different, complex compartments such as quarantined [19], hospitalized [20], diagnosed [18] among others.

B. Asymtomatic infection and SIR-type Models

In medicine, a asymptomatic patient is one that tests positive for a disease but shows no symptoms [19]. Asymptomatic carriers can shed the disease to those around them, but generally, at a slower rate than those that are symptomatic [3]. The main issue is that asymptomatic carriers do not know they have the disease; thus, they may not follow the same procedures as someone who knows they are infectious would. For example, someone who has a cough, generally, will try and cover their mouth to protect those around them, if they did not have any noticeable symptoms, they will spread the disease unknowingly [7].

The asymptomatic effect has caused problems in disease tracking and planning for many diseases including COVID-19 [4]–[6]. The proportion of asymptomatic infections that make up the COVID-19 pandemic has been widely debated, and it may been anywhere from 4%-80% [3]–[7]. The problem limiting these studies is how the authors validate and reliably collect data. Studies which focus on asymptomatic have the difficulty of finding these infections as the carriers are not evaluated. With the possibility of having a significant amount of COVID-19 asymptomatic carriers shedding the disease to those around them, it is crucial to understand how much of an impact they are having on overall case counts. The next challenge is tracking the impact of asymptomatic cases on the overall true case count. Tracking these cases and seeing who they infect is even harder. This leaves modelers with the job of estimating how many asymptomatic carriers are in each population and how many people they are exposing.

There have been different studies on the integration of the asymptomatic state in disease spread modelling. One of these studies [20] proposed a SIARD (Susceptible, Infected, Asymptomatic, Recovered, Dead) and a SQIARD model (where Q is the Quarantine state). The SIARD model uses a simple transition from the susceptible state to the infectious or asymptomatic state using a given infectious rate. The SQIARD model incorporated the asymptomatic state as a transition from the quarantine state. The model splits the population that moves from the quarantine state to the asymptomatic or infectious state using specific rates. The model provides results that resemble real world case counts using different countries. Another study [21] proposed an advanced, SIDARTHE model, which also includes states for diagnosed (D), ailing (A), recognized (R), threatened (T), healed (H) and extinct (E) individuals. Asymptomatic individuals are added by using multiple disease subcategories under one state: asymptomatic is a subcategory of the infected state. When an individual moves from susceptible to infected they can become asymptomatic, infected, or undetected. The asymptomatic individuals will then move to either diagnosed, ailing or healing. If an asymptomatic individual becomes detected, they are considered diagnosed asymptomatic. Asymptomatic individuals who move to ailing develop symptoms and become undiagnosed symptomatic, and those who move to healing will recover from the infection. The proportion of individuals who move to each state is defined by the states specific transition rate, i.e., those who move from asymptomatic infected to ailing is denoted by the probability for a host to develop symptoms. The authors also remark the importance of those who are asymptomatic or undetected as they will not be isolating like those who are known infectious.

In [18], the authors present a SEAIRD model that showed a similar transition method as those described previously. They use an asymptomatic state where an asymptomatic rate α is defined to split the infected population into infectious or

asymptomatic. Susceptible individuals can become exposed to the virus (E) or remain in the susceptible (S) state. Exposed individuals can become asymptomatic (A = αE) or infectious (I = α (1- E)). The model showed how asymptomatic cases can affect the rate and magnitude of new infectious cases in a population. The authors also described a more advanced model called the SEAIRD-Control model where a quarantine state, and a hospitalization state have been introduced.

Although our focus is on the influence of asymptomatic cases, these extra states (quarantined, hospitalized, etc.) can be considered as future additions to our model. None of the models above include the geographical aspect of the disease where the relationship between two neighborhoods impacts how the disease can spread. The model we are proposing in this paper addresses this idea and incorporates geographical attributes in disease transmission dynamics.

C. SIR-type Models including geographical aspects

Another advancement made in SIR modelling was the addition of geographical information. Sattenspiel and Dietz [10] describe a model for the spread of infectious diseases among geographic regions. They describe how individuals can become mobile and be in contact with individuals in other regions, resulting in the spread of a disease across regions. They show how geographical information can complement the standard SIR model as well as lead to better, more defined results. In [22], the authors describe a geographical Cell-DEVS SIR model. Their model is based on [23] to simulate the spread of epidemics in a geographical based 2D cell space. The model has described that at time *t*, a given cell (i, j) has a given population N_{ij}. Each cell stores the ratio of individuals in each state. The SIR model described in [23] and translated to Cell-DEVS in [22] uses a geographical correlation factor defined by the shared boundaries between two cells. The correlation factor is a method the model uses to link two regions together to allow for interaction between their populations. This is not necessarily the most accurate method as it does not take other key factors into account. For example, it does not consider dense population areas and workplace hubs such as downtown Ottawa. Adding transportation networks [24] or human movement and mixing models [25] showing how a population move from one region to another would give more accurate results. However, the method has been shown to produce accurate results and it is easily adaptable for geographical simulations [23]. Being able to quickly adapt a model to receive new data is crucial for rapid prototyping, and this model allows user to change both the geographical level they are simulating as well as the disease characteristics. The equations for the correlation factor between neighborhoods (i.e., cells) are as follows:

$$w_{ij} = w_{ji} = \frac{\frac{z_{ij}}{l_i} + \frac{z_{ji}}{l_j}}{2} \quad (2.1)$$

$$c_{ij} = w_{ij} \quad (2.2)$$

Equation (2.1) describes the weighted correlation factor w_{ij}, which uses the two values, the shared boundary length between cells *i* and *j* (Z_{ij}, Z_{ji}) in both directions, divided by the total boundary length of cells *i and j* (l_i, l_j). This method states that the correlation for *i*, when moving to *j*, is the same as *j* moving to *i*. Finally, equation (2.2) is used to set the geographical correlation factor between cell *i* and *j*, this will be used in Section III. The model developed in [22] also includes parameters that define hospital capacities and lockdowns correction factors.

In [22], the authors extend the geographical model described in [23] to incorporate deaths and the ability for a cell's recovered populations to become re-infected [26][27]. In [28], the authors extend the model with the ability for a cell's population to move from the susceptible state to an exposed state before becoming infected, developing the SEIRD model presented that was used as starting point for the SEAIRD model framework we propose in this research. When adding the exposed state, the rate at which a cell's population would move from susceptible to exposed remained unchanged. But, when the population moved from susceptible to exposed, a new value was considered, the incubation rate, ε. The exposed state is an important addition to the model as it allows for incubation rate simulations to be added. The model can simulate realistic calculations showing the time it takes for someone to be exposed to when they show symptoms. The problem with this is that not all individuals show symptoms, thus the need for an asymptomatic stage. Our main contribution is the addition of asymptomatic individuals in a geographical SIR-type model.

The models discussed in section II.B define different methods that can be incorporated into a geographical SEAIRD model. Each model shares a defined asymptomatic rate where a given proportion of the exposed population move to either infectious, or asymptomatic. We follow a method that allows for easy, efficient implementation where user can prototype different values without having to change more than the input parameters. To do so, we use a spatial modeling methodology, called Cell-DEVS [12].

D. Cell-DEVS and Cadmium

Cell-DEVS [12] is a modeling methodology that allows defining cell spaces based on the Discrete Event Systems Specifications (DEVS) [29]. Cell-DEVS describes an n-dimensional cell space where each cell represents a DEVS atomic model. The cell space containing the *n* cells is defined as a DEVS coupled model where each cell is connected to its neighboring cells, as in Figure 1.

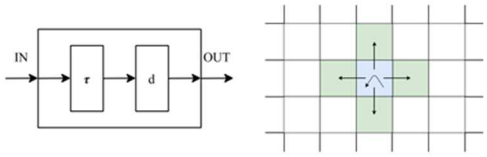

Fig. 1. Cell-DEVS model: (a) Atomic cell schematics; (b) 2-dimensional Cell-DEVS neighborhood

When a cell receives an input, the local computing function τ is activated, this will compute the next state for the cell. This discrete-event approach only considers and computes active cells using a continuous time base. If there is a change in the cell's state, the change is transmitted after a time delay *d*. In figure 1(b), we can see how a cell (center) will transmit information to the neighboring cells using a von

Neumann neighborhood. Cell-DEVS accepts other neighborhoods and irregular topologies as well. Cell-DEVS inherits the modularity and hierarchical modeling ability of DEVS. This allows for models to better interact with other models, tools, datasets, and visualization tools, making it an easy, and efficient method to build complex cellular models.

There are different simulators to execute Cell-DEVS models [13]. In this research, we use the Cadmium tool [13], which allows users to define model inputs using JavaScript Object Notation (JSON), a data format to store and transmit large amounts of human readable data. JSON stores data in key-value pairs allowing for the simple representation of neighborhoods, their attributes, and their relationships. Cadmium allows the user to include complex geographical inputs that load into the model at run time resulting in a flexible model that allows for efficient rapid prototyping.

III. Model Specifications

Our proposed geographical SEAIRD model is based on [22], [28] by adding an asymptomatic state (A) as depicted in the diagram in Figure 2, which shows that a cell's population starts in a susceptible state and then it can become exposed. From there, the exposed population will move to either asymptomatic or infectious. If asymptomatic, they will eventually become recovered, but if an individual is infectious, they can move to either recovered or deceased.

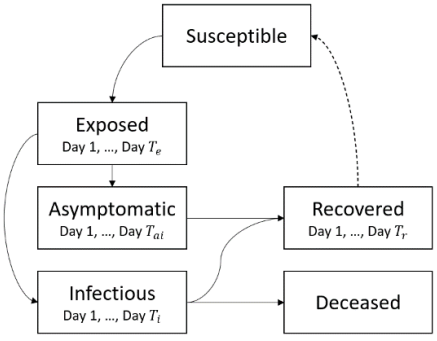

Fig. 2. SEAIRD State Diagram

The dotted line in Figure 2 from recovered to susceptible shows how a population can become re-susceptible after recovery. Each transition is based on their defined time behavior and described using the delay function. Exposed, Infectious, Asymptomatic and Recovered states have a defined set of days that a population can be within the state described by T_e, T_i, T_{ai}, and T_r. Each state has a defined state transition that occurs at each day within the state. The days within each state set of days is described by $q = \{1, 2, \ldots, T_{state}\}$. For example, $A_{i,a}^t(q)$, describes the proportion of asymptomatic cases for an age group in cell i at asymptomatic state $q = \{1, 2, \ldots, T_{ai}\}$, at time t. The asymptomatic state works similarly to the infectious state: an asymptomatic rate describes the proportion of the population that transitions from exposed to either infectious or asymptomatic.

Our model uses k unique geographical cells. The proportion of a population's age group a found in each state is described by: $S_{i,a}^t, E_{i,a}^t, A_{i,a}^t, I_{i,a}^t, R_{i,a}^t, D_{i,a}^t$, where i is the cell being described at time t. The state transitions are built using the Cell-DEVS transition and delay functions, which implement equations 3.1 – 3.12.

Let us consider $fa(q)$ as the fatality rate of infected stage q for age group a; $\lambda a(q)$ their virulence; $\mu a(q)$ their mobility rate; $\varepsilon a(q)$ the incubation rate; $\gamma a(q)$ the recovery rate and φ as the asymptomatic infection rate. Then, c_{ij} is the geographical correlation factor between cells i and j; k_{ij} is the correction factor applied to both cells i and j to model disobedience.

$$D_{i,a}^{t+1} = D_{i,a}^t + \sum_{q=1}^{Ti} fa(q)\left(I_{i,a}^t(q)\right) \tag{3.1}$$

Eq (3.1) is used to calculate the proportion of deaths at a time t. The proportion of deaths next day is the total of current deaths plus the sum of the infectious population that died the day before. New deaths are equal to the newly deceased population moving from the infectious state multiplied by the fatality rate. The deceased transition does not consider asymptomatic infections as they do not lead to deaths.

$$E_{i,a}^{t+1}(1) = S_{i,a}^t \sum_{j=1}^{k} \begin{pmatrix} c_{ij}k_{ij} * \sum_{\substack{b\in A \\ p\in\{1,2,\ldots,T_i\}}} \frac{N_{j,b}}{N_j} \mu_b(p)\lambda_b(p) I_{j,b}^t(p) \end{pmatrix} + \left(c_{ij} * \sum_{\substack{b\in A \\ p\in\{1,2,\ldots,T_i\}}} \frac{N_{j,b}}{N_j} \mu_b(p)\lambda_b(p) Ai_{j,b}^t(p) \right) \tag{3.2}$$

Equation (3.2) is used to calculate the proportion of newly exposed population. This is a result of the susceptible ones in contact with either the entire infectious population or the asymptomatic population of neighboring cells j. The first part of the equation calculates the proportion of a cell's susceptible population exposed to an infectious individual (I) and the second part the proportion exposed to an asymptomatic individual (Ai). A defines the set of age groups in cell j, each age group is represented by b. Each cell's population represented by N_j is divided into age groups (the subscript b). Each cell is related to its neighbor by a geographical correlation factor c_{ij} that describes the impact each neighboring cell has on a given cell, including virulence and mobility rates a given cell's population has with its neighbors. Finally, k_{ij} defines a correction factor between cells i and j, applied to the infectious half of the equation to simulate different behavior for infectious and asymptomatic populations: we consider that asymptomatic individuals to be more carefree, thus they will expose more individuals. The correction factor k_{ij} is defined using the models disobedience factor d where $k_{ij}=min(k_i, k_j)$. The correction for individual cells i and j is defined as $k_{cell}=d+(1-d)*m_c$. The infection correction factor mc is defined in the model as a function of the infection threshold (I_{TH}) that triggers a specific mobility correction factor (cm) and a hysteresis level (H).

$$E_{i,a}^{t+1}(q) = \left(1 - \varepsilon_a(q-1)\right) E_{i,a}^{t+1}(q-1) \quad q \in (2, 3, \ldots, T_e) \tag{3.3}$$

Equation (3.3) describes how the exposed population transitions to the infectious or asymptomatic state. The equation defines the exposed in stage q is equal to the exposed of the previous day multiplied by $1- \varepsilon_a(q-1)$. Where $\varepsilon_a (q-1)$ defines the incubation rate for an age group a for state $q - 1$. The incubation rate defines the probability of the population moving to infectious or asymptomatic.

$$I_{i,a}^{t+1}(1) = E_{i,a}^{t}(T_e) + \sum_{q=1}^{Te-1} \left(\varepsilon_a(q) E_{i,a}^{t}(q)\right)(1-\varphi) \quad (3.4)$$

Equation (3.4) describes the new infectious population that will occupy day 1. The equation considers the exposed population from all stages, and all age groups. As defined above in (3.3), a proportion of the exposed population moves to infectious or asymptomatic depending on the incubation rate ε_a. The rate at which the exposed population becomes either infectious, or asymptomatic is defined by asymptomatic rate φ. Thus, for the case of new infectious population the rate is defined as (1 – φ).

$$I_{i,a}^{t+1}(q) = I_{i,a}^{t}(q-1) * \big(1 - \gamma_a(q-1) - fa(q-1)\big) \quad (3.5)$$
$$q \in (2, 3, \dots, T_i)$$

Equation (3.5) describes the portion of the infected population that moves to the next stage. The infectious population for stage q equals the population of infectious in the previous stage, $q - 1$ minus the population who move to either recovery or deceased. The portion of the population that move to the recovered or deceased states is defined by recovery rate γ and fatality rate f respectively.

$$A_{i,a}^{t+1}(1) = E_{i,a}^{t}(T_e) + \sum_{q \in \{1,2,\dots,T_e-1\}} (\varepsilon_a(q) E_{i,a}^{t}(q))\varphi \quad (3.6)$$

$$A_{i,a}^{t+1}(q) = A_{i,a}^{t}(q-1) * \big(1 - \gamma_a(q-1)\big) \quad (3.7)$$
$$q \in (2, 3, \dots, T_{ai})$$

Equations (3.6 and 3.7) define the asymptomatic state behavior following the same rules described in (3.4) and (3.5). Equation (3.6) defines the proportion of the exposed population that moves to the asymptomatic state (here, the asymptomatic population rate remains as φ). Equation (3.7) follows the same rules defined when asymptomatic cases either move to the next stage q, recovered, or deceased.

$$R_{i,a}^{t+1}(1) = \begin{array}{c} \left(I_{i,a}^{t}(T_i) + A_{i,a}^{t}(T_{ai})\right) + \\ \sum_{q \in \{1,2,\dots,T_i-1\}} \gamma_a(q) I_{i,a}^{t}(q) + \\ \sum_{q \in \{1,2,\dots,T_{ai}-1\}} \gamma_a(q) A_{i,a}^{t}(q) \end{array} \quad (3.8)$$

Equation (3.8) describes the proportion of infectious or asymptomatic populations that become recovered. The equation defines that the total number of recoveries is equal to the total number of recoveries from the previous day plus the newly recovered population. The current day recoveries are calculated by taking the proportion of infectious and asymptomatic infections that move to the recovered stage using rate γ. Finally, the equation checks for the population that is on the final day of either infectious or asymptomatic, if their population does not move to the deceased state, they are added to the recovered state.

$$R_{i,a}^{t+1}(q) = R_{i,a}^{t}(q-1) \quad (3.9)$$
$$q \in \{2,3, \dots, T_r - 1\}$$

$$R_{i,a}^{t+1}(T_r) = R_{i,a}^{t}(T_r) + R_{i,a}^{t}(T_r - 1) \quad (3.10)$$

Equations (3.9 and 3.10) are used only if re-susceptibility is not enabled. Once the recovered population reaches the final day of recovery, they remain there for the rest of the simulation time.

$$R_{i,a}^{t+1}(q) = R_{i,a}^{t}(q-1) \quad (3.11)$$
$$q \in \{2,3, \dots, T_r\}$$

Equation (3.11) is an equation only used when re-susceptibility is enabled, i.e., patients who are recovered will go through each day of recovery, when they reach the final day of recovery the population will move back into the susceptible population pool where they can be re-exposed.

$$S_{i,a}^{t+1} = \begin{array}{c} 1 - \sum_{q=1}^{Te} E_{i,a}^{t+1}(q) - \sum_{q=1}^{Ti} I_{i,a}^{t+1}(q) \\ - \sum_{q=1}^{Tai} A_{i,a}^{t+1}(q) - \sum_{q=1}^{Tr} R_{i,a}^{t+1}(q) \\ -D_{i,a}^{t+1} \end{array} \quad (3.12)$$

Equation (3.12) is a "special equation" needed for the integrity of the model. Since we know that any given population starts in the susceptible state (excluding the starting cell) then the population that is not in any other state should remain susceptible.

The model is defined as a coupled Cell-DEVS where the cell space represents a geographical region, and each cell (of irregular topology) is a district in the city/province. It relates to its neighboring cells using an irregular topology. Each cell consists of a cell ID, a set of state variables, a model configuration, and neighboring cell's correlation factors.

IV. Model Implementation

We implemented the equations defined in section III, and when all the geographical cells are defined, they are placed into a top level coupled cell model called geographical_coupled, with configuration seen in Figure 3.

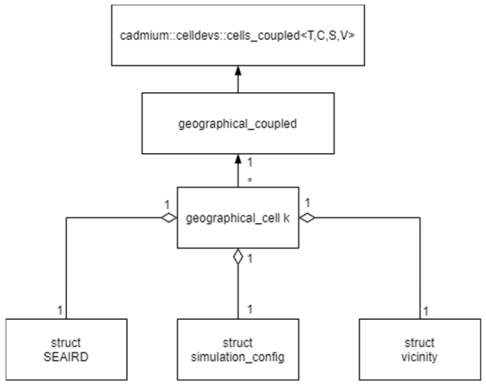

Fig. 3. SEAIRD Coupled Cell Diagram

At runtime, the geographical_coupled model is initialized using cell's data provided from in a JSON input file (using the methods described in the top model class cadmium::celldevs:cells_coupled<T,C,S,V>; Figure 3 shows how this coupled cell model is defined. At the bottom level of figure 3 the three structures define the inputs to the geographical cells. The SEAIRD structure defined the state variables that will hold the population as well as the infection correction factors and the disobedience factor (Figure 4). The simulation configuration structure defines the attributes used to characterize the disease being modelled including recovery

rates, fatality rates and asymptomatic rates (Figure 5). The vicinity structure holds the information that defines the correlation factors between two cells (geographical_cell). The three structures are read in at run time to create the single parameterized model geographical_cell. The collection of geographical cells and their relationships define the geographical coupled model.

```
struct seaird {
    std::vector<double> age_group_proportions;
    std::vector<double> susceptible;
    std::vector<std::vector<double>> exposed;
    std::vector<std::vector<double>> infected;
    std::vector<std::vector<double>> asymptomatic;
    std::vector<std::vector<double>> recovered;
    std::vector<double> fatalities;
    std::unordered_map<std::string,hysteresis_fact
    or> hysteresis_factors;
    double population;
    std::vector<double> disobedient; ...};
```

Fig. 4. SEAIRD configuration code

Each cell contains the relevant information defined in the SEAIRD configuration file. At run time, each cell has a unique population which is divided into described age groups. Each cell's population will then be divided into one of the six states. If modelling the beginning of a pandemic, a single cell will hold the initial case(s) and the remaining cells will be 100% susceptible. At t=0 the proportion of a cell's population in each state is defined in by the values provided in the SEAIRD structure. Defining the SEAIRD structure at run time with user defined inputs allows users to choose the point of time they want to start a model (if they are interested in the middle of a pandemic they can tailor the input values to hold the number of individuals in each state at that time).

```
struct simulation_config {
    int prec_divider;
    using phase_rates =
    std::vector<std::vector<double>>;
    phase_rates virulence_rates;
    phase_rates incubation_rates;
    phase_rates recovery_rates;
    phase_rates mobility_rates;
    phase_rates fatality_rates;
    double asymptomatic_rates;
    bool SIIRS_model = true;
};
```

Fig. 5. Simulation configuration

In figure 5 the simulation configuration is declared; these values are used to change the ways a population transferred from one state to another. These structures define the geographical_cell atomic model presented in Figure 3. A geographical_cell atomic model is defined for each geographical cell in the model, these cells make up a geographical_coupled model where each cell is connected by a correlation factor. Finally, this geographical_coupled model is defined by methods found in class cadmium::celldevs::cells_coupled<T,C,S,V>.

V. Case Study: Ontario Public Health

The SEAIRD results presented in this section are generated using source data from the 34 Ontario public health units where the population is generated using census data [30]. A configuration file is built using a geo package (Geopandas [31]) to determine shared boundaries (correlation factor) between each public health unit. Figures 6-9 were created using the graphing tools in [22], R [32] and plotly [33]. The implementation of the model is available at `https://github.com/SimulationEverywhere-Models/Geography-Based-SEAIRDS`. The results presented in this section show and compare the effect the asymptomatic state has when added to the simulation. The parameters used in this study are shown in Table 1:

TABLE I. Test case configuration

Parameter	Value
Population	Varies per cell based on census data [34]
Age Groups	[0.216, 0.279, 0.268, 0.193, 0.044] [34]
Disobedience	[0.29, 0.25, 0.23, 0.21, 0.24] [35], [36]
Asymp. Rate	Varies per simulation (See figure 6-8)
Virulence	0.1 across all states and age groups
Incubation	14-day profile [28]
Mobility Rates	1.0 across all states and age groups
Recovery Rates	0.07 across all states and age groups
Fatality Rates	0.005 across all states and age groups
Infection correction factors (lockdown)	0.001:[0.60, 0.0008], 0.005:[0.50, 0.003], 0.01: [0.40, 0.005], 0.03: [0.30, 0.015], 0.08: [0.20, 0.0005], 0.15: [0.1, 0.08], 0.20: [0.01, 0.12]

We start with the population of each geographical area. We then use a vector of values to represent the proportion of the total population in each age group. Age groups are an abstract set of values defined by the modeler; in our case, individuals 0-12, 13-19; 20-44; 45-65, and over 65 years old. Disobedience follows the same format as age groups where the values in the vector represent the proportion of each age group that is disobedient to lockdowns. These values were estimated using data gathered from [35], [36]. The asymptomatic rate is the proportion of the exposed population that become asymptomatic (the rest become infectious). The virulence rate represents the rate at which the disease spreads, the value represents the amount of age groups population in contact with the infected population per day of the infection. Incubation rate is the proportion of the exposed population that become infectious or asymptomatic, defined using a 14-day profile where each day a proportion of the exposed population will move to the next state [28]. Mobility rates define the freedom of the population to move (1.0 mobility rating means the population can move freely). Mobility rates are defined for each age group for each day of the infection, it is assumed that mobility rates are not restricted at all at the beginning of the pandemic. Recovery rates define the proportion of an age group infected population that will recover each day of the infection. Fatality rates defines the proportion of an age group infected population that will move to the deceased state on each day. Infection correction factors describe the proportion of the population that needs to be infected before a lockdown is put in place. The values can be described as follows: 'Proportion of population infected to start lockdown': ['mobility modifier', 'Proportion of population infected to life lockdown'] where mobility modifier reduces the mobility of a cell by the given value.

Virulence rates, recovery rates, fatality rates and infection correction factors were informed by data gathered at [2]. The values were evaluated and slightly modified; the final values shown in Table 1 provided the most accurate results in testing.

Figure 6 shows the simulation results with 0% asymptomatic cases. The results show the steady rise of exposed individuals (orange line), then 1-14 days after their exposer they become infected (red line). Our results show the initial wave rises and settles in a little over 350 days with approximately 8% of the population becoming infected.

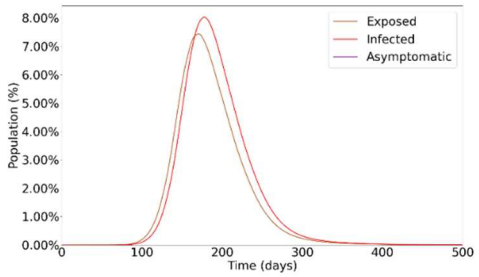

Fig. 6. SEAIRD Model - 0% Asymptomatic

Next, we studied the effect of an 80% asymptomatic rate using the same parameters. Figure 7 shows a lower rate of infectious carriers and a higher rate of asymptomatic carriers.

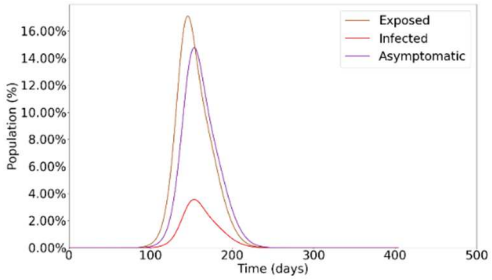

Fig. 7. SEAIRD Model - 80% asymptomatic

With this higher rate of asymptomatic carriers, the total exposed population reaches a higher peak than it had without asymptomatic cases (approximately 110% more exposures occur). This higher exposed count is due to the different asymptomatic and infectious carriers have on the susceptible population. Since the asymptomatic carriers travel more than the infectious carriers more of the population is exposed to them, causing higher overall infections. The initial curve rises and settles in approximately 250 days this is 100 days less that the model with no asymptomatic infections. This shows that the asymptomatic carriers expose the susceptible population at a much higher rate than the model showing no asymptomatic cases. We can see approximately 14% of the population become asymptomatic carriers with an additional 4% being infectious carriers. Although in this case we can see higher overall rates of infections, the asymptomatic infections are less lethal, leading to less deaths and more recoveries. We see this difference when analyzing a single neighborhood cell in figures 8 and 9.

Figure 8 shows a single cell and how its population transitions through the states with a 0% asymptomatic rate. It should be noted the cell is still exposed to its neighboring cells. When examining the graph at day 133 we can see a cumulative exposed population of 7.1%, cumulative infectious population of 5.65%, cumulative fatalities of 0.4%, and since we have no asymptomatic infections, an asymptomatic infected population of 0%.

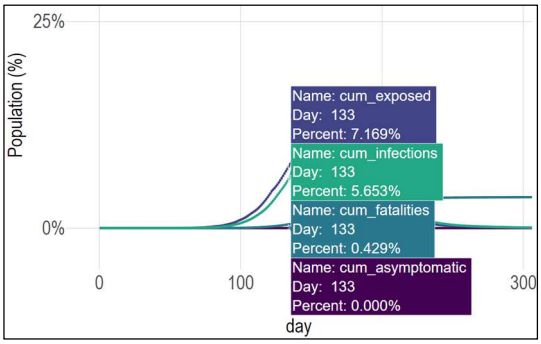

Fig. 8. Single Cell SEAIRD - 0% Asymptomatic

If we then compare this to a graph showing the curves with an asymptomatic rate of 80% (Figure 9), we can see that at day 133 we have a cumulative exposed population of 27.7%, cumulative infectious population of 4.7%, cumulative fatalities of 0.02% and an asymptomatic population of 19.4%. We can see that the exposed grown to 27.7%, from 7.1%. This 20% increase can be accredited to the asymptomatic carriers spreading the disease to the surrounding neighbors at a higher rate than the infectious population. We can also see that we have fewer deaths, linked to the fewer infectious cases.

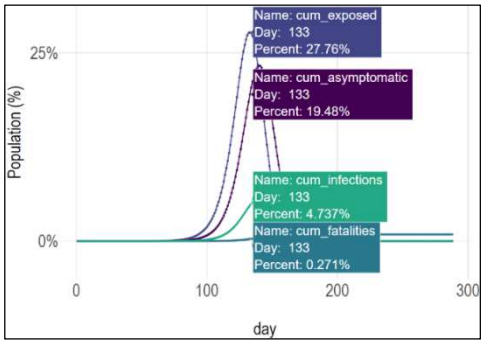

Fig. 9. Single Cell SEAIRD - 80% Asymptomatic

We can now simulate the effect a “invisible” population of disease carriers may have on a pandemic combined with the advantage of having a model that includes geographical aspects. If 80% of total COVID-19 cases are asymptomatic and the case counts show 5% of a population are infected, we can expect that 20% more of the population is also infected but not showing symptoms and not aware they are infected. These asymptomatic individuals will be spreading the disease to the healthy population, causing rises in the number of exposed individuals, resulting in more infectious. Although a high asymptomatic rate will lead to more overall cases, these cases are not as deadly, this is due to asymptomatic cases not showing symptoms and in the case of COVID-19 not leading to death. If we were to model a disease where asymptomatic cases could lay dormant for years and later lead to death we would see more interesting fatality results.

VI. Conclusions

We presented a model that allows users to create rapid simulation prototypes to simulate how much of an impact asymptomatic cases would have on disease case counts. Another important note about the model, although it has been built around the COVID-19 pandemic, it can be used for any other disease. This can be done in a quick, efficient manner. If users have the relevant information for a disease along with the asymptomatic rate at which a disease transmits, they can simply change the parameters and re-run the model. The model also gives the users the ability to efficiently adapt the geographical level that the model is being run on.

Future adaptations plan to incorporate asymptomatic disease transmission rates (modified virulence rates), asymptomatic rates by age group, a more accurate susceptible to exposed transition and new compartments to represent individuals that are vaccinated, and therefore less susceptible to the disease. The model can be easily adapted to simulate variants of concern by tuning the test case configuration data found in Table 1. In future adaptations variants of concern will be addressed in a formal manner allowing for the simulation of single and multiple variants of concern.

Acknowledgements

This project is funded by NSERC Alliance research grants.

References

[1] World Health Organization, "WHO Coronavirus (COVID-19) Dashboard," 2021. https://covid19.who.int/.

[2] Public Health Ontario, "Ontario COVID-19 Data Tool," 2021. https://www.publichealthontario.ca/.

[3] M. Yanes-Lane et al., "Proportion of asymptomatic infection among COVID-19 positive persons and their transmission potential: A systematic review and meta-analysis," PLoS One, vol. 15, no. 11, pp. 1–21, 2020.

[4] D. Buitrago-Garcia et al., "Occurrence and transmission potential of asymptomatic and presymptomatic SARSCoV-2 infections: A living systematic review and meta-analysis," PLoS Med., vol. 17, no. 9, pp. 1–25, 2020.

[5] O. Byambasuren, M. Cardona, K. Bell, J. Clark, M. L. McLaws, and P. Glasziou, "Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis," J. Assoc. Med. Microbiol. Infect. Dis. Canada, vol. 5, no. 4, pp. 223–234, 2020.

[6] A. M. Pollock and J. Lancaster, "Asymptomatic transmission of covid-19," BMJ, vol. 371, pp. 19–20, 2020.

[7] M. Ralli, A. Morrone, A. Arcangeli, and L. Ercoli, "Asymptomatic patients as a source of transmission of COVID-19 in homeless shelters," Int. J. Infect. Dis., vol. 103, pp. 243–245, 2021.

[8] D. P. Oran and E. J. Topol, "Prevalence of Asymptomatic SARS-CoV-2 Infection : A Narrative Review," Ann. Intern. Med., 2020,

[9] Flaxman, S., Mishra, S., Gandy, A. et al. "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe." Nature, vol. 584, pp. 257–261, 2020.

[10] L. Sattenspiel and K. Dietz, "A structured epidemic model incorporating geographic mobility among regions," Math. Biosci., vol. 128, no. 1–2, pp. 71–91, 1995.

[11] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," Proc. R. Soc. Lond., no. A 115, pp. 700–721, 1927.

[12] G.A. Wainer, "Cellular Modeling with Cell-DEVS: A Discrete-Event Cellular Automata Formalism," in Cellular Automata, 2019, pp. 6–15.

[13] L Belloli, D. Vicino, C. Ruiz-Martin, and G.A. Wainer, "Building Devs Models with the Cadmium Tool", in Proceedings - Winter Simulation Conference, National Harbor, 2019, pp. 45–59.

[14] R. Ross, "An application of the theory of probabilities to the study of a priori pathometry.—Part I," Proc. R. Soc. Lond., vol. A 92, pp. 204–230, 1916.

[15] R. Ross and H. P. Hudson, "An application of the theory of probabilities to the study of a priori pathometry.—Part II," Proc. R. Soc. Lond., vol. A 93, pp. 212–225, 1917.

[16] R. Ross and H. P. Hudson, "An application of the theory of probabilities to the study of a priori pathometry.—Part II," Proc. R. Soc. Lond., vol. A 93, pp. 212–225, 1917.

[17] H. W. Hethcote, "Mathematics of infectious diseases," SIAM Rev., vol. 42, no. 4, pp. 599–653, 2000.

[18] M. Barman, S. Nayak, M. K. Yadav, S. Raha, and N. Mishra, "Modeling control, lockdown & exit strategies for COVID-19 pandemic in India," medRxiv, July, 2020.

[19] B. C. Lin, Y. J. Chen, Y. C. Hung, C. S. Chen, H. C. Wang, and J. L. Chern, "The data forecast in COVID-19 model with applications to US, South Korea, Brazil, India, Russia and Italy," arXiv, pp. 1–15, 2020.

[20] G. Giordano et al., "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy," vol. 26, no. 6. Springer US, 2020.

[21] I. Potasman, "Asymptomatic Infections: The Hidden Epidemic," Int. J. Clin. Res. Trials, vol. 2, no. 2, 2017.

[22] R. Cárdenas, K. Henares, C. Ruiz-Martín, and G. Wainer, "Cell-DEVS Models for the Spread of COVID-19," 14th International Conference on Cellular Automata for Research and Industry, Lodz, Poland, 2020.

[23] S. B. Zhong, Q. Y. Huang, and D. J. Song, "Simulation of the spread of infectious diseases in a geographical environment," Sci. China, Ser. D Earth Sci., vol. 52, no. 4, pp. 550–561, 2009.

[24] Li, Jian et al. "Modeling epidemic spread in transportation networks: A review." Journal of Traffic and Transportation Engineering, 2021.

[25] Kraemer, M.U.G., Golding, N., Bisanzio, D. et al. "Utilizing general human movement models to predict the spread of emerging infectious diseases" in resource poor settings. Sci Rep 9, 5151, 2019.

[26] Ota, M. "Will we see protection or reinfection in COVID-19?". Nat Rev Immunol 20, 351, 2020.

[27] Stokel-Walker C. "What we know about covid-19 reinfection so far" BMJ 2021; 372 :n99.

[28] G. Davidson, G. Wainer, "Studying Covid-19 Spread Using a Geography Based Cellular Model". Winter Simulation Conference. Phoenix, AZ. 2021

[29] B. Zeigler, H. Praehofer, and T. G. Kim, "Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems". San Diego, CA: Academic Press, 2000.

[30] Government of Ontario - Ministry of Health, "Public Health Unit Locations". 2021. https://www.health.gov.on.ca/en/common/system/services/phu/locations.aspx

[31] Kelsey Jordahl, Joris Van den Bossche, Martin Fleischmann, Jacob Wasserman, James McBride, Jeffrey Gerard, François Leblanc. geopandas/geopandas: v0.8.1. Zenodo.

[32] Team R Development Core, "A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, 2018. http://www.r-project.org.

[33] Plotly Technologies Inc., "Collaborative data science," Plotly Technologies Inc., 2015. https://plot.ly.

[34] Statistics Canada. Census of Canada: Profile Data for Ottawa. Ottawa, Ont, 2016

[35] L. Carlucci, I. D'ambrosio, and M. Balsamo, "Demographic and attitudinal factors of adherence to quarantine guidelines during covid-19: The italian model," Front. Psychol., 2020.

[36] Angus Reid Institute, "COVID-19 Compliance: One-in-fiveCanadians making little to no effort to stop coronavirus spread," 2020,http://angusreid.org/covid-compliance/.