Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo and M. Loper, eds.

GEOGRAPHY BASED CELLULAR COVID-19 DISEASE SPREAD MODEL

Glenn Davidson
Gabriel Wainer

Department of Systems and Computer Engineering
Carleton University
1125 Colonel By Drive
Ottawa, ON K1S 5B6, CANADA

ABSTRACT

Infectious disease models are in widespread use among governments and health agencies to plan COVID-19 public health policies, and their accuracy is of upmost importance to public health. Common approaches to modeling infectious diseases include compartmental differential equation models and Cellular Automata, both of which are difficult to use in predicting the spread of disease over several geographical regions, as they often oversimplify the geographical features they are meant to model. A geography-based Cell-DEVS approach to modelling pandemics is presented. The compartmental model presented considers additional factors such as movement restriction effects, disease incubation, population disobedience to public health policies, and a dynamic fatality rate. The model offers deterministic predictions for any number of regions simultaneously and can be easily adapted to unique geographical areas.

1 INTRODUCTION

In December 2019, a novel coronavirus named SARS-COV-2 emerged and became a global pandemic over the span of one month. Actions were taken by governments to restrict international travel and encourage the social isolation of their citizens, but all precautions proved insufficient, and the virus persists globally as of April 2021. The cumulative global fatality count due to COVID-19 infection exceeds 3.1 million people as of May 2021 (Johns Hopkins Coronavirus Resource Center 2021).

To forecast and prevent the spread of infectious diseases, organizations and governments use mathematical models to make informed policy decisions. Such models have been used in the past to predict the spread of Ebola and HIV (Rachah and Torres 2016; Djordjevic et al. 2018). A popular method of modelling infectious disease spread is to use compartmental differential equation models, describing how individuals in different states of infection interact and influence one another. Often missing from this approach is the consideration of how populations are distributed over physical space, and how they travel between unique geographical regions. Instead, Cellular Automata (CA) models can begin to address the spatial distribution of disease by dividing a region under study into cells and applying differential equation relationships between neighboring cells. This division is traditionally a 2D uniform square grid, where cells are related to one another in a uniform neighborhood pattern. However, the geography in which populations reside is rarely uniform, contradictory to the common use of regular neighborhood patterns in CA.

Considering these issues, we defined a Cell-DEVS COVID-19 model to predict the spread of COVID-19 in distinct geographical areas using a geography-based model. Cellular approaches to the modelling of geographical areas have been studied, but few models consider the irregular correlation between geographical areas, and even fewer have applied cellular geography in the modelling of disease spread (Zhong et al. 2009). Section 2 gives the necessary background information of the model. Section 3 presents a formal model definition of the Geographical SEIRD COVID-19 model and provides a description of the Cadmium Cell-DEVS implementation of the model. Section 4 describes how the geography of Ontario was used to simulate the model, and how COVID-19 case data from the Public Health Units of Ontario were used to calibrate the model's parameters.

2 BACKGROUND

Mathematical disease models date back to the 17th century, where Bernoulli published the first model to assess the benefits of population variolation to smallpox (Dietz 2002). In 1927 (Kermack et al. 1927) defined the Susceptible, Infected, and Recovered (SIR) model using differential equations. This model remains the basis of many infectious disease spread models used at present because of its simplicity and success in describing the spread of disease. The model presented in this paper is an SIR based model.

2.1 SIR Models

The SIR model divides each member of a population into one of three categories: Susceptible (S), Infected (I), and Recovered (R), and uses ordinary differential equations to describe how individuals transition between categories. Susceptible people are vulnerable to the pathogen and will become infected if they have sufficient contact with infected people, which spread the disease to Susceptible people through their daily contacts according to a virulence rate of the disease (λ). Infected people have a certain probability of recovering each day, defined as the recovery rate (γ). Recovered people have overcome the disease, are no longer infectious, and have become permanently immune. A set of differential equations can be derived to describe how a population's daily interactions cause infection and immunity to disease over time:

$$\begin{cases} S = S(t) & s(t) = \frac{S(t)}{N} \\ I = I(t) & i(t) = \frac{I(t)}{N} \\ R = R(t) & r(t) = \frac{R(t)}{N} \end{cases} \quad (1) \qquad \begin{cases} \frac{ds}{dt} = -\lambda s(t) i(t) \\ \frac{dr}{dt} = \gamma i(t) \\ \frac{di}{dt} = \lambda s(t) i(t) - \gamma i(t) \\ s(t) + i(t) + r(t) = 1 \end{cases} \quad (2)$$

The set of equations (1) defines the percentage of the population in each of the infection states over time as the total number with the infection state at time t, divided by the size of the population N. The set of equations (2) define how Susceptible individuals are infected by their contacts with the Infected individuals proportional to λ. The recovery rate γ defines the rate at which Infected individuals become Recovered each day. The final part of the equation states that the sum of all individuals in all infection states is always equal to one, implying no birth, death, or migration. The behavior of all individuals is also assumed to be identical. These non-idealities are sources of error for which the model's results will deviate from real pandemics because they do not account for individualistic human behavior and the variance in disease effects between individuals. The trajectory of a SIR model is predictable: a population begins as nearly 100% Susceptible, Infections rise and then fall, increasing the number of Recovered over time:

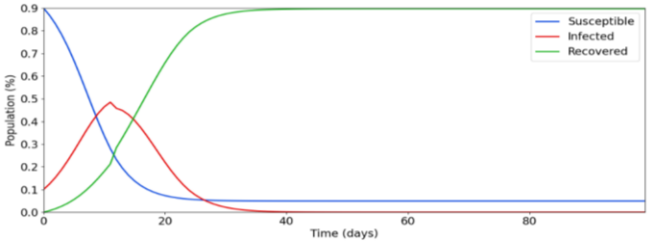

Figure 1: SIR Model Trajectory

To increase the accuracy of the SIR model, other compartmental states have been used, including an Exposed state (E) that describe the latency period preceding contagious infection, Quarantined or Isolated states (Q) which describe a period of reduced human contacts while in a stage of infection, Asymptomatic states (A) which defines individuals who are not aware of their infection but are contagious, and a Dead/Fatal (D) state. Many combinations and variations of the states described have been researched for the purposes of modelling infectious diseases. The spread of COVID-19 using non-geographical SIR models have successfully predicted outbreaks in individual cities and countries. (Caccavo 2020) describes

a SIRD model that correctly estimates the spread of COVID-19 in China and Italy. A SEIIR model presented by (Danon et al. 2020) predicts the COVID-19 outbreaks and Wales and England.

A variety of CA models have also been developed to study the spatial spread of COVID-19. (Medrek and Pastuszak 2021) describe a CA SEIR model that uses the age group distributions and population counts of regions to construct a heterogeneous cell space. The model includes geographical considerations in creating cell spaces but uses a regular neighborhood pattern. A probabilistic SEIQR CA defined in (Ghosh and Bhattacharya 2020) use similar methods of population density and variation to define a cell space representing a country. A genetic algorithm is used in selecting the disease parameters based on what is known from the epidemiological data of near areas. The research in (Tobler 1975) gives a foundation of cellular geography, including the description of geographical regions as cells in both regular and irregular neighborhood patterns. (Zhong et al. 2009) define a geographical SIRS CA using irregular geography-based cell spaces, where a cell's neighborhood is defined by the amount of border length shared with other regions. This neighborhood definition allows cells to have unequal influences on one another, preserving the shape of geographical regions within the model.

The Discrete Event System Specification (DEVS) is a formalism used to model and simulate discrete-event systems. The DEVS framework has been used to implement the rules cell spaces as a discrete-event models, giving the Cell-DEVS formalism described in (Wainer 2014). Cell-DEVS can be used to specify and implement cellular models and facilitates their simulation and integration with other models. Cell-DEVS is advantageous in that no consideration of time steps or synchronization delays are required, as inputs to cells are processed immediately, and their outputs are set after a delay function. (Cárdenas et al. 2020) introduced a SIR model built in Cell-DEVS using the C++ Cadmium DEVS simulation library.

2.2 Geographical Cellular Models

Uniform neighborhood patterns may not always best represent cellular models of geographical regions. For some applications it may be better to use a neighborhood definition that allows cells to have differing numbers of neighbors, size and weights of influence. In a cellular geographical model, the behavior of a cell's local transition function should vary depending on the shape and size of the cell and nearby cells. Neighborhoods may be described in terms of correlational weights between regions instead of simply placing cells in one another's neighborhoods based on adjacency. Tobler's first law of geography states that in the context of geography, "everything is related to everything else, but near things are more related than distant things". This law serves as a heuristic in determining the level of geographical correlation between areas (Waters 2017). Considering a cell space formed from the 2d map of Figure 2, variation can be seen in the distances between region centroids and in the amount of border length shared between regions.

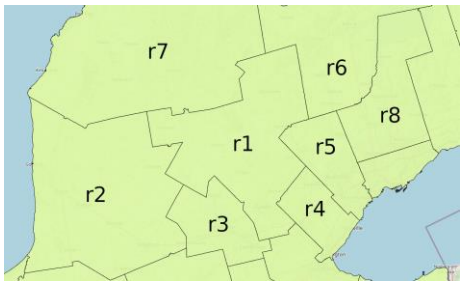

Figure 2: Geographical Cells

A simple method to determine geographical neighborhoods is to consider all cells in a cell space as related by a correlational weight between each pair of regions. The neighborhood pattern can be described as: every cell is included within every other cell's neighborhood with a correlational weight. However, if the correlational weight between two regions is zero, their relationship need not be considered. The method of geographical correlation we used considers the amount of shared border length between regions. To

determine the proportion of shared border length z_{ij} between two regions i and j with respect to the total length of border of each region can be calculated as z_{ij} / li for cell i, and z_{ij} / lj for cell j. This results in different w_{ij} when correlating cell i to cell j, and cell j to cell i, so an average of two factors is taken.

2.3 SEIRD Cellular Automata

A CA SEIRD model predicts the state of infection of a group of cells over time, where the state variables S, E, I, R, D describe the percentage of total cell population having that state. The states Exposed, Infected, and Recovered are best described as consisting of multiple phases, each corresponding to a single day. The Susceptible State describes individuals who have not yet encountered the disease and will become Exposed if a sufficient contact is made with an Infected person. The Exposed state describes individuals who have the disease but have not yet developed symptoms and are not yet contagious. The length of the exposed stated is T_E days. The Infected state describes individuals who have caught the disease and are contagious, lasting for T_I days. The Recovered state describes individuals who have overcome the disease and are no longer contagious, lasting for T_R days. In a model with reinfection, the Recovered have immunity from disease for a determinate amount of time before becoming Susceptible again. If reinfection is not modeled, then the Recovered state can be represented in a single phase. The Fatal state describes fatalities that result from infection, having no interaction with the population, but remaining in the population count.

$$
\begin{cases}
E_{i,a}^{t+1}(1) = S_{i,a}^{t} \sum_{j \in V} c_{ij} * \sum_{\substack{b \in A \\ p \in \{1,2,\dots,T_i\}}} \frac{N_{j,b}}{N_j} \mu_b(p) \lambda_b(p) I_{j,b}^{t}(p) & (a) \\
E_{i,a}^{t+1}(q) = \big(1 - \varepsilon_a(q-1)\big) E_{i,a}^{t}(q-1) & (b) \\
\qquad q \in \{2, 3, \dots, T_e\} & \\
I_{i,a}^{t+1}(1) = E_{i,a}^{t}(T_e) + \sum_{q \in \{1,2,\dots T_e - 1\}} \varepsilon_a(q)\, E_{i,a}^{t}(q) & (c) \\
I_{i,a}^{t+1}(p) = I_{i,a}^{t}(p-1)(1 - \gamma_a(p-1) - \digamma_a(p-1)) & (d) \\
\qquad p \in \{2, 3, \dots, T_i\} & \\
R_{i,a}^{t+1} = R_{i,a}^{t} + I_{i,a}^{t}(T_i) * (1 - \digamma_a(T_i)) & (e) \\
\qquad + \sum_{p \in \{1,2,\dots,T_i - 1\}} \gamma_a(p)\, I_{i,a}^{t}(p) & \\
D_{i,a}^{t+1} = D_{i,a}^{t} + \sum_{p \in \{1,2,\dots,T_i\}} \digamma_a(p)\, I_{i,a}^{t}(p) & (f) \\
S_{i,a}^{t+1} = 1 - \sum_{q \in \{1,2,\dots,T_E\}} E_{i,a}^{t+1}(q) - \sum_{p \in \{1,2,\dots,T_i\}} I_{i,a}^{t+1}(p) & (g) \\
\qquad - R_{i,a}^{t+1} - F_{i,a}^{t+1} &
\end{cases}
\tag{3}
$$

The cell space of the SEIRD CA model is formed using 2 dimensional geography based neighborhoods, consisting of a set of cells M = (m_1, m_2,...,m_i) each with population Ni. Each cell has a unique neighborhood definition V_i that is defined as a set of cell names (neighbors) and each cell's correlational weight, V_i = {(m_i, 1), (m_k, c_{ik}),...,(m_x, c_{ix})}. The population in a specific phase q of Exposed can be further specified as $E_i^t(q)$. Similarly, the phases of Infected are specified using the index p. Each cell population N_i is divided into age groups, specified by the subscript a, which describes the proportion of total cell population N_i that are members of age group a. Dividing each population into age groups allows the incorporation of the age specific factors that largely determine human behavior and disease outcomes. The probability that an Exposed individual enters the Infected state is controlled by the incubation rate ε(q). The probability that an Infected individual becomes Recovered is controlled by the recovery rate γ(p). The probability that an Infected individual dies and enters Fatal from each phase of Infected is controlled by the fatality rate ϝ(p). The contagiousness of individuals in each phase of Infected is given by the virulence rate λ(p), which approximates the epidemiological R_o of the disease over the phases of Infected. The mobility rate of individuals is described the mobility rate μ(p).

The local transition function of the discrete time SEIRD model is defined by equations 3a-3g. The capital A symbol in 3a is defined as the set of age groups in cell j. The $c_{ij} \in [0,1]$ is the geographical correlation factor between regions i and j and describes the amount of population interaction between cells i and j. Equation (3a) describes how the Susceptible population of cell i ($S_{i,a}^{t}$) interacts with neighboring Infected populations (I_j^t) of all age groups, resulting in new Exposed ($E_{i,a}^{t+1}$). The first Exposed phase population (new Exposed) is equal to the Susceptible population multiplied by the Infected in each phase p, the virulence of phase p, the mobility of infected individuals in phase p, and c_{ij}. Equation (3b) describes the population in all Exposed phases except the first. The population in all subsequent Exposed phases is the Exposed population in the previous phase multiplied by 1 minus the incubation rate of that phase. Equation (3c) describes the newly infectious population in Infected phase 1. The population in Infected phase 1 is equal to the population leaving the Exposed state, where all individuals that reach the final Exposed phase become Infected on the next time step. Equation (3d) states that the Infected population in each subsequent phase of Infected is equal to the Infected of the previous phase, minus the recoveries and fatalities that occur during that phase. Equation (3e) states that the Recovered population is equal to the Recovered population of the previous time step, plus the sum of all new recoveries from each phase of Infected, plus the entirety of all Infected that reach phase T_i and do not die. The Fatal equation (3e) states that the total fatalities within a cell are equal to the fatalities of the previous time step plus all new fatalities occurring due to each phase of Infected. The Susceptible Equation states that the sum of all infection states within an age group is always equal to one. The sum of all age group proportions in A is also always one.

3 MODEL DEFINITION

The SEIRD model defined in section 2 was used to define a Cadmium Cell-DEVS model. The default model parameters used in experimentation were defined using known disease parameters of COVID-19. The mean incubation rate of COVID-19 is 5.4 days following a log-normal distribution (McAloon et al. 2020). This distribution is used to define both T_E=14 and the values of incubation rate $\varepsilon(q)$. The mean symptomatic infection length is 10 days, which was used to develop an infection profile using T_I =12 days, where the recovery rate $\gamma(p)$ controls the proportion that recovered in each phase (CDC 2021). The virulence and fatality rates were varied to match infection case data in experimentation. The fatality rate in each age group in a cell depends on the cumulative sum of infections across all age groups of that cell. If the level of infection within a cell grows beyond a threshold, the fatality rate increases according to a fatality rate modifier F. Several new variables are introduced to define an extended correlation factor g_{ij}, called the infection correction factor, which is directly substituted for c_{ij} in the local transition rules defined in section 2. A movement restriction factor k_i is introduced to define movement restriction policies, and their effectiveness in reducing disease spread. Several movement restriction factors may be assigned to a single cell to define tiers of lockdown intensity. When calculating the total infection correction factor between two cells, the more severe k_i in effect is used to describe their interaction. The geographical correlation and neighborhood definition remains the same, and the rules of equations 3a-3g still apply.

The infection threshold defines the cumulative amount of active infection above which the mobility rate multiplier μk will be active. The mobility rate multiplier of the k_i becomes inactive when the level of active infections falls below the infection threshold level minus the hysteresis level. Only a single k_i may be active at any time, and the model will select the most restrictive applicable k_i depending on the current level of infection. The infection correction factor g_{ij} can be calculated as a function of the geographical correlation factor, the mobility rate multiplier, and the disobedience factor d which describes the proportion of total population that follows the k_i movement restrictions. When considering the interaction between two cells, the more restrictive infection correction factor is used. The more severe a lockdown policy is, the lower the movement restriction multiplier is, reducing g_{ij} and the infective correlation between the cells.

The model is implemented as a Cadmium Cell-DEVS model "geographical_coupled", consisting of atomic cell models of type "geographical_cell" that are coupled to form a cell space. Each geographical_cell uses string cell id (C), a seird state object that contains all cell state variables (S), and a

vicinity struct that describes a cell's neighborhood (V). The relationships between the classes involved can be seen in the Unified Modeling Language (UML) diagrams of Figures 4 and 5:

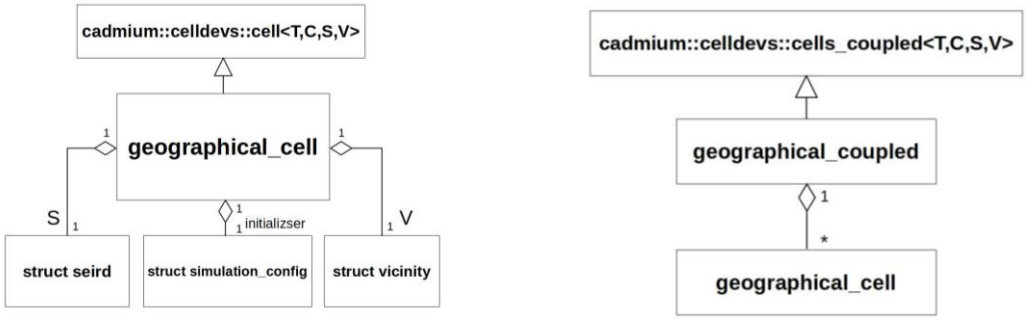

Figure 4: Geographical Cell model UML Class Diagram

The infection state variables computed in the local transition function are discretized per equations 11a to 11e, where P is the precision divider, and the use of square brackets denotes a round operation:

$$\begin{cases} DS_{i,a}^{t} = \frac{[PS_{i,a}^{t}]}{P} & (a) \\ DE_{i,a}^{t}(q) = \frac{[PE_{i,a}^{t}(q)]}{P} & (b) \\ \quad q \in [1, 2, \dots, T_e] & \\ DI_{i,a}^{t}(p) = \frac{[PI_{i,a}^{t}(p)]}{P} & (c) \\ \quad p \in [1, 2, \dots, T_i] & \\ DR_{i,a}^{t} = \frac{[PR_{i,a}^{t}]}{P} & (d) \\ DF_{i,a}^{t} = \frac{[PF_{i,a}^{t}]}{P} & (e) \end{cases} \tag{4}$$

The local computation function implemented in the geographical_cell class is described by the pseudocode of Table 1. The algorithm begins by allocating a state object new in line 1, which will be used to store the new values of cell state variables. Lines 2 through 8 calculate the new state variables for each age group based on the cell's current state object s and stores the result in new. These six lines implement equations 3a-3g, and the discretization in equations 4a-4e are used. The resulting state object new is returned in line 9 and will be used elsewhere to write the new state to the calling geographical_cell.

The method new_exposed() described in the local computation function is responsible for calculating the infective interactions between all neighboring cells and is the most complex portion of the local computation function. The calculation of new exposures can be explained by the pseudocode in Table 2

Table 1: Geographical Cell local computation function

Input: none. **Output:** New State (struct seird)

```
1: new = seird()
2: for(each age group a in s.age_groups)
3:   new.E.at(a) = new_exposed(a,s)
4:   new.I.at(a) = new_infected(a,s)
5:   new.R.at(a) = new_recovered(a,s)
6:   new.D.at(a) = new_fatal(a,s)
7:   new.S.at(a) = 1 - new.E.at(a) - new.I.at(a) - new.R.at(a) - new.D.at(a)
8: endfor
9: return new;
```

The calculation of new exposures begins by calculating the movement correction multiplier of cell i, using the method movement_correction() which scans the set K_i and chooses the most suitable movement restriction factor to apply. The infection correction factor g_i is calculated in line 2 by considering the proportion of disobedient di that will follow the movement restriction policy. In line 3 the variable infStrength is declared and will be used to sum how much infective contact is made with cell i from each of its neighbors. A for loop is used in lines 4-16 to iterate over all neighbors of cell i, summing the amount of infective with all neighbors. This is done by first calculating the infection correction factor of cell j, and then choosing the more restrictive correction factor between i and j to describe their interaction (lines 5-7).

Table 2: New Exposures Algorithm

Input: age segment *b*, current state *si*. **Output:** New Exposed

```
1: µi = movement_correction(Ki, infected_i)
2: gi = di+(1-di)*µi
3: infStrength = 0
4: for(all neighbors j of cell i)
5:   µj = movement_correction(Kj , infected_j)
6:   gj = dj +(1-dj)* µj
7:   gij =  min(gi, gj)
8:   for (all age groups a in cell j)
9:     iSG = 0;
10:    for (q in infected phases)
11:       iSG += sj.λ(a,q) * sj.µ(a,q) * sj.I(a,q)
12:    endfor
13:    gij = kij*cij
14:    infStrength += (Nj.at(a)/Nj)*iSG*gij
15: endfor
16: endfor
17: new_exposed = si.S.at(b)* infStrength
18: return new_exposed
```

Next the algorithm calculates the infective potential of each age group a in cell j by summing over all infection phases q the proportion of population with phase q and their infective strength in that phase (line 11). The total infective weight of the age group is then calculated in line 14 as a function of the proportion of population in that age group, the infective potential of the age group, the infection correction factor k_{ij}, and the geo-correlation factor c_{ij}. The algorithm concludes in lines 17 and 18 by multiplying the Susceptible population in age group b by the sum of infective strength of the neighborhood and returning the result.

4 EXPERIMENTAL RESULTS

The map of Figure 6 (OHESI 2021) shows the province of Ontario, Canada, including the 34 Public Health Units (PHUs) of Ontario we used in testing the model defined in section 3.

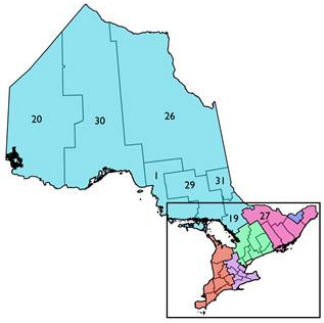

Figure 6: PHUs of Ontario (OHESI 2021)

The simulation results were compared to the publicly reported case data in these regions for the period of January 1st, 2020 to February 2nd, 2021. The case data used was reported by the Government of Ontario, in the form of confirmed new cases per date and their resulting outcome (Government of Ontario 2021). The SEIRD model predicts the number of active cases but does not report the data in the form of new cases per day. To calibrate and validate the model results, the confirmed cases were converted to active infections and cumulative fatalities over time. Each confirmed case was assumed to follow the mean infection length of COVID-19 when converting case reports to active cases. A single confirmed case is converted into 10 days of active infection, starting on the reported case date. Fatal outcomes were still considered to produce contagious individuals with an active infection length of 10 days. The geographical boundary file used in defining the cell space for the Ontario Public Health units (Ontario GeoHub 2021).

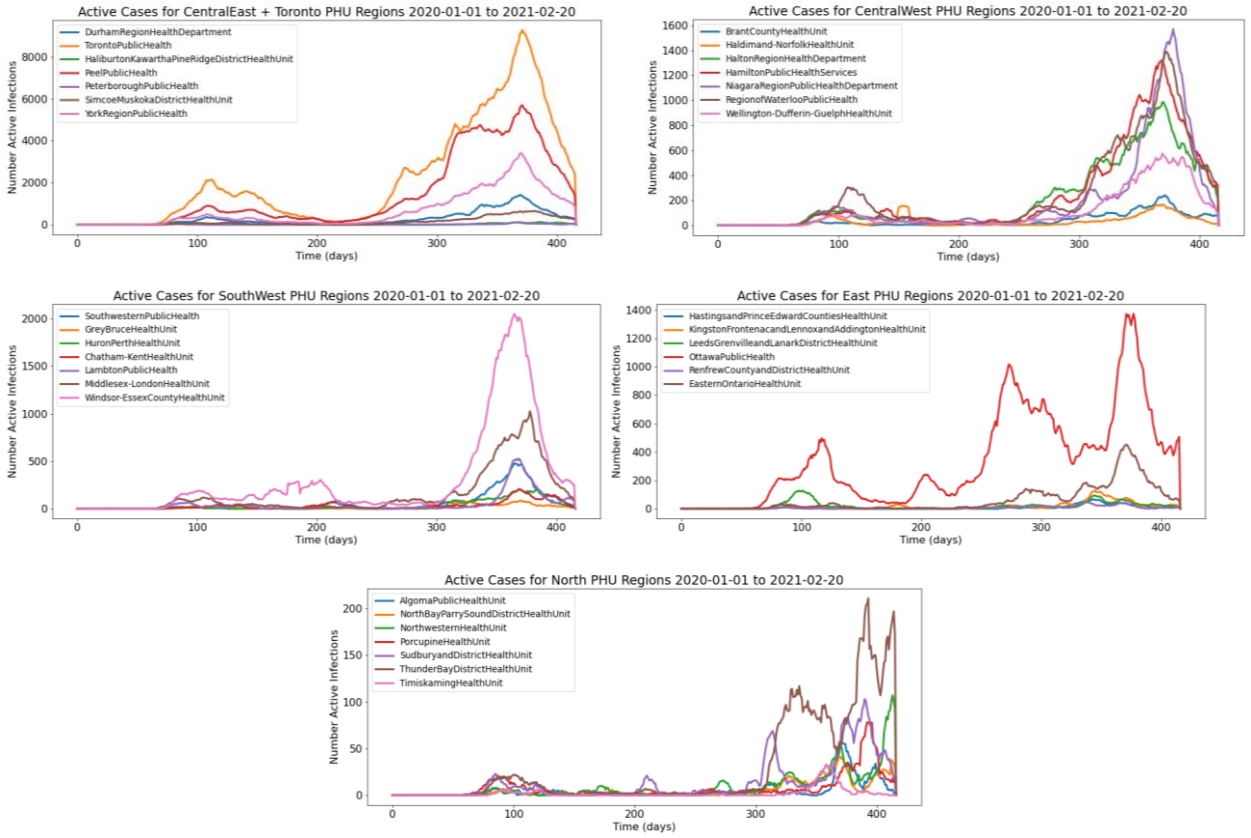

Figure 7: Ontario PHU Case Data from Dec 2020 to February 2021: (a) Central East PHUs; (b) Central West PHUs; (c) South West PHUs; (d) East PHUs; (e) North PHUs.

The PHUs of Ontario were placed in 5 groups based on their location: Central East, Central West, South West, East, and North. A visualization of the case data over for these groups is shown in Figure 7.

Table 3: Default Virulence Rates and Infection Correction Factors

Parameter:	**Value:**
virulence_rate	0.07 all days
Ki {}	"0.0005": [0.80, 0.0003]
	"0.001": [0.40, 0.0009]
	"0.01": [0.30, 0.009]
	"0.05": [0.10, 0.04]
	"0.10": [0.05, 0.08]
	"0.15": [0.00, 0.12]

The default experimentation parameters and the structure of the state variables used in experimentation can be found in https://github.com/SimulationEverywhere. All default values are used in the following tests unless explicitly redefined. The default values of the virulence rate and K_i are given in Table 3.

Without movement restriction policies to limit infection spread, most of the population in all 34 cells encounter the disease in a sever trajectory like Figure 8. The infection consists of one wave in which over 90% of the population becomes Infected, and the majority become Recovered according to the fatality rate.

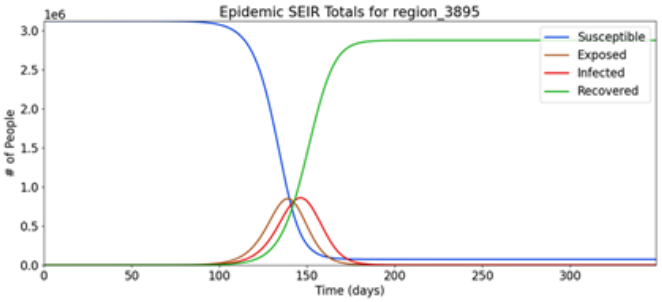

Figure 8: Toronto Pandemic, No Movement Restriction

We examined the effect of movement restriction policies in the Toronto PHU in terms of the movement correction factor's infection threshold, hysteresis, and mobility multiplier. Figure 9 shows simulation data compared to the reported active cases. We used an infection threshold too high, resulting in movement restrictions that are applied later than what is seen in the case data. The predicted active infections would have been more accurate had the infection threshold of the limiting policy been invoked as 50% Infected.

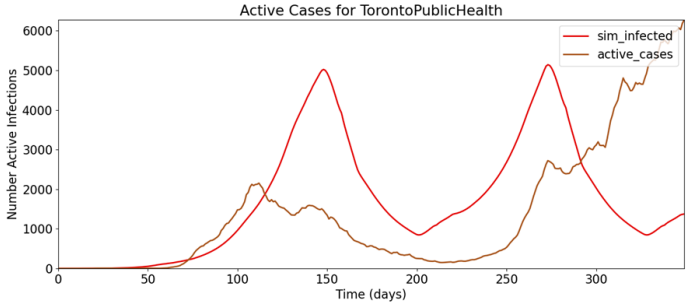

Figure 9: Toronto Pandemic, Movement Restriction Factor #1

By using an infection threshold 50% lower than the simulation of Figure 9, the movement restriction policy comes into effect at a more accurate time, as shown in Figure 10.

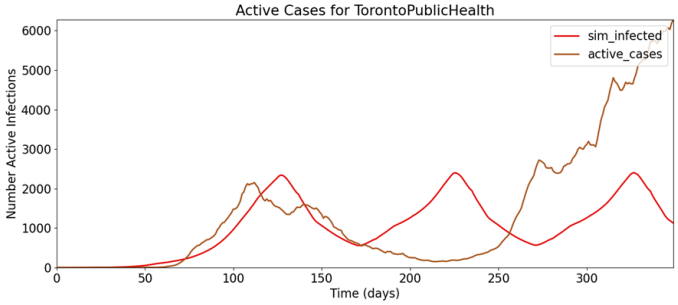

Figure 10: Toronto Pandemic, Movement Restriction Factor #2

However, these restrictions are lifted too early, and a second wave begins, while the case data indicates a continual fall of cases at this time. This indicates that not enough hysteresis was used in the infection correction, and that the hysteresis factor should be increased such that the number of active cases will fall to a lower level before movement restrictions are lifted. A larger hysteresis factor results in a better match of the Toronto PHU infection, shown in Figure 11:

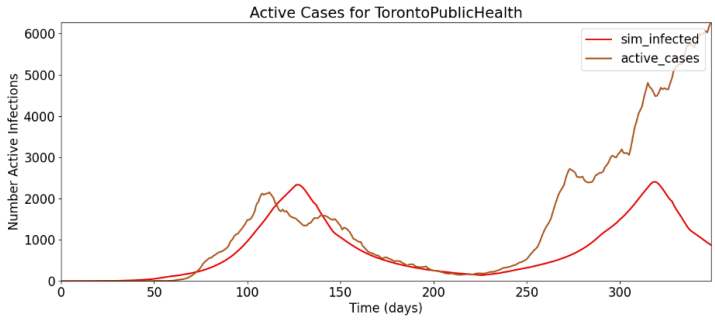

Figure 11: Toronto Pandemic, Movement Restriction Factor #3

After these preliminary tests, we conducted a series of simulations and compared to the known case data to calibrate the parameters of the model. The virulence rate and infection correction factors were found to be the most influential determinants of model accuracy, which was improved by increasing virulence rate above the default level, and the overall strictness of lockdown policies, given in Table #4.

Table 4: Best Resulting Virulence Rates and Infection Correction Factors

Parameter:	**Value:**
virulence_rate	0.30 all days
Ki {}	"0.00002": [0.50, 0.00001]
	"0.0003": [0.40, 0.00025]
	"0.0007": [0.15, 0.00065]
	"0.02": [0.10, 0.01]
	"0.05": [0.02, 0.025]
	"0.10": [0.00, 0.06]

While the first wave of the pandemic was strongest in Toronto, other regions considerably further away also had significant levels of infection relative to their total population size, namely the Ottawa and Windsor Essex PHUs. A single infection source beginning in the Toronto cell was insufficient in producing the first wave infection curves far away from Toronto, so a smaller amount of infection was added to the initial conditions of Ottawa and Windsor Essex. These two areas also happen to be a major border crossing area into regions outside of the cell space: Quebec in the case of Ottawa, and the USA in the case of Windsor Essex. Figure 12 shows the results in regions surrounding Toronto, Ottawa, and Windsor Essex.

The simulations in these regions do track the general shape of the case data, but in some cases there is significant overshoot in case estimation. Toronto has the highest overall accuracy as the model parameters were initially developed to match Toronto before considering more regions. Identical infection threshold rules are being applied uniformly to each cell; they are dependent on the percentage of a cell's total population that is infected. This could be a source of error because factors such varying human behavior and population density from cell to cell are not considered in the lockdown policy thresholds. In fact, in the case of the regions adjacent to Toronto, the population density of each region relative to Toronto's population density is proportional to error in simulated cases versus reported cases. The population density of the Peel PHU is 25% that of Toronto. Similarly, the population density of York is 11.74% of Toronto, and Durham is 5.46% of Toronto. It may not be a coincidence that Durham, having a the least similar population density to Toronto, is the neighbor of Toronto with highest the highest inaccuracy.

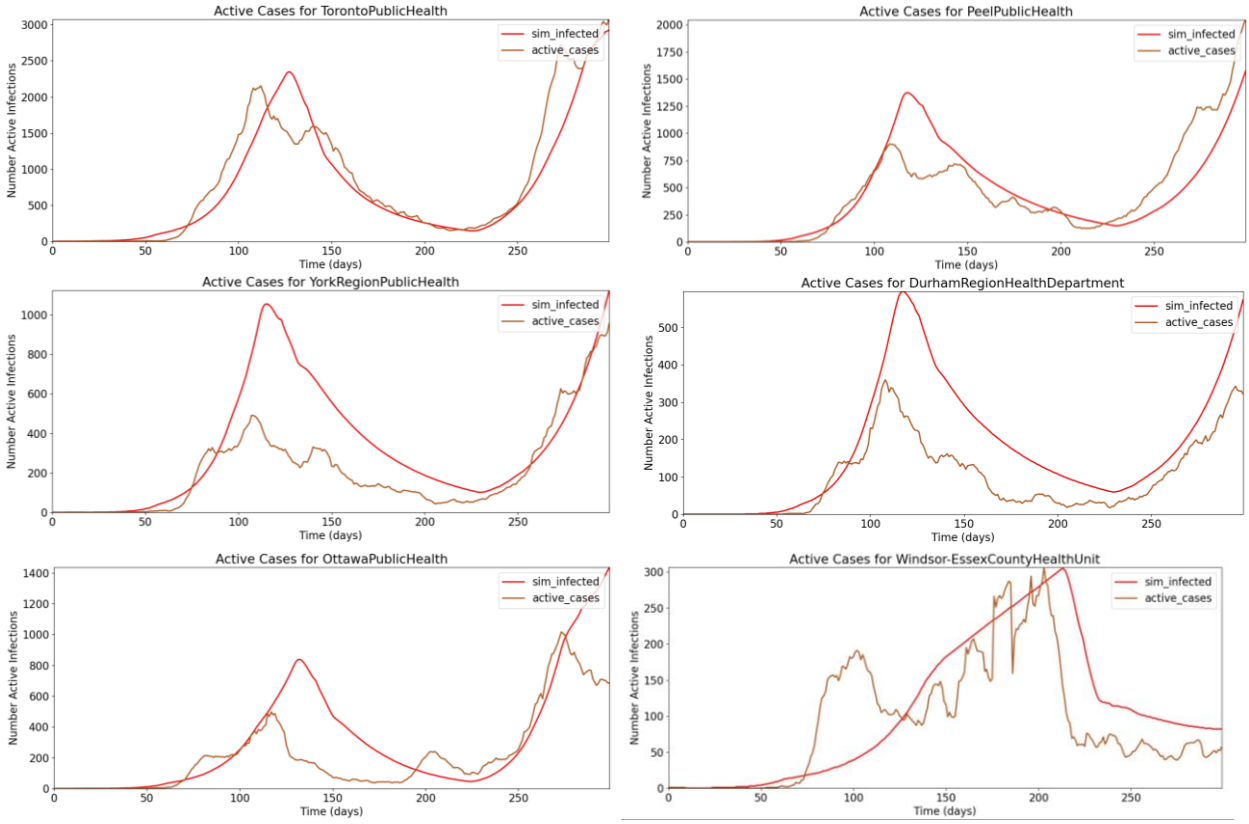

Figure 12: Pandemic in (a) Toronto; (b) Peel; (c) York; (d) Durham; (e) Ottawa; (f) Windsor Essex.

A GIS view of these simulation results showing how the beginning of the first wave progresses can be seen in Figure 13. Note that some infection spreads in Ottawa before Toronto's wave has a large influence. Although Ottawa and Windsor Essex are initialized with similar amounts on infection, the infection does not spread as fast in Windsor Essex because it shares less border length with adjacent regions than Ottawa.

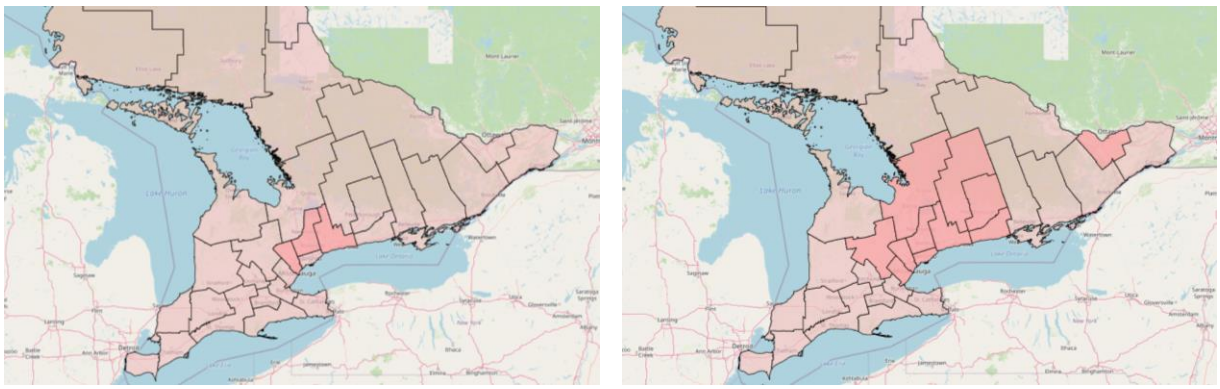

Figure 13: First Wave Pandemic at (a) 88 days; (b) 95 days;

5 CONCLUSIONS

A Cell-DEVS COVID-19 model was defined and simulated to study the spread of infectious disease in existing geographical environments. The model defined in this paper uses a method of geographical CA like that used by Zhong et al in their 2009 simulation of the original SARS pandemic, but considers more

factors in the implementation of movement restriction policies. The model was simulated using the geography of Ontario, Canada, and divided the province into geographical cells by Public Health Unit. The methodology of adjusting the model and disease parameters was demonstrated, and the accuracy of the model was improved substantially. Many directions for future improvements exist including:

- The addition of a different class of cell to represent the coupling of each cell to the outside world.
- The addition of a population density correction factor.
- Increasing disobedience or changing the infection correction factors over time.
- Examining results further away from the infection origin to determine their source of error.
- Searching literature on the effect of seasons in disease spread as the impact could be significant in areas that experience freezing winters.

REFERENCES

Caccavo, D. 2020. “Chinese and Italian COVID-19 Outbreaks can be Correctly Described by a Modified SIRD Model. *MedRxiv*.

Cárdenas, R., K. Henares, C. Ruiz-Martín, and G. Wainer. 2021. “Cell-DEVS Models for the Spread of COVID-19”. *Proceedings of the 2020 International Conference on Cellular Automata for Research and Industry*, December 2nd–4th, Łódź, Poland.

CDC. 2021. “Interim Guidance on Ending Isolation and Precautions for Adults with COVID-19”. https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html, accessed 6th April.

Danon, L., E. Brooks-Pollock, M. Bailey, M. Keeling. 2020. “A Spatial Model of CoVID-19 Transmission in England and Wales: Early Spread and Peak Timing”. *MedRxiv*.

Dietz, K., J. Heesterbeek. 2002. “Daniel Bernoulli's Epidemiological Model Revisited” *Mathematical Biosciences* 180(1-2):1-21.

Djordjevic, J., C. J. Silva, and D. F. M. Torres. 2018. “A Stochastic SICA Epidemic Model for HIV Transmission”. *Applied Mathematics Letters* 84:168-175.

Government of Ontario. 2021. “Confirmed Positive Cases of COVID-19 in Ontario”. Ontario Open Data Catalog, https://data.ontario.ca/en/dataset/confirmed-positive-cases-of-covid-19-in-ontario, accessed 2nd February.

Ghosh, S., and S. Bhattacharya. 2020. “A Data-Driven Understanding of COVID-19 Dynamics using Sequential Genetic Algorithm Based Probabilistic Cellular Automata”. *Applied Soft Computing* 96:106692.

John Hopkins Coronavirus Resource Center. 2021. “COVID-19 Map”. https://coronavirus.jhu.edu/map.html, accessed 3rd May.

Kermack, W.O., A. McKendrick, and G. T. Walker. 1927. “A Contribution to the Mathematical Theory of Epidemics”. *Proceedings of the Royal Society of London A* 115(772):700–721.

McAloon, C., et al. 2020. “Incubation Period of COVID-19: A Rapid Systematic Review and Meta-Analysis of Observational Research”. *BMJ Open* 10(8):e039652.

Medrek, M., and Z. Pastuszak. 2021. “Numerical Simulation of the Novel Coronavirus Spreading”. *Expert Systems With Applications* 166:114109.

Ontario GeoHub. 2021. “Ministry of Health Public Health Unit Boundary”. https://geohub.lio.gov.on.ca/datasets/ministry-of-health-public-health-unit-boundary, accessed 2nd February.

OHESI. 2021. “Health Regions”. http://www.ohesi.ca/health-regions, accessed 3rd May.

Rachah, A., and D. F. M. Torres. 2016. “Dynamics and Optimal Control of Ebola Transmission”. *Mathematics in Computer Science* 10(3):331-342.

Tobler, W. R. 1979. “Cellular Geography”. In *Philosophy in Geography*, edited by G. Olsson, 379–386. Dordrecht: Springer.

Wainer, G. A. 2014. “Cellular Modeling with Cell-DEVS: A Discrete-Event Cellular Automata Formalism”. In Proceedings of the 2014 International Conference on Cellular Automata for Research and Industry, September 22nd-25th, Kraków, Poland.

Waters, N. 2018. “Tobler’s First Law of Geography”. In *International Encyclopedia of Geography*, edited by D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu and R. A. Marston. Hoboken, New Jersey: Wiley.

Zhong, S., Q. Huang, D. Song. 2009. “Simulation of the Spread of Infectious Diseases in a Geographical Environment”. Science in China Series D: Earth Sciences 52(4):550-561.

AUTHOR BIOGRAPHIES

GLENN DAVIDSON is a master’s student at the Department of Systems and Computer Engineering at Carleton University. His email address is glenn.davidson@carleton.ca

GABRIEL A. WAINER is Professor at the Department of Systems and Computer Engineering at Carleton University (Ottawa, ON, Canada). He is a Fellow of SCS. His email address is gwainer@sce.carleton.ca