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

# STUDYING COVID-19 SPREAD USING A GEOGRAPHY BASED CELLULAR MODEL

Glenn Davidson
Gabriel Wainer

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

## ABSTRACT

The study of infectious disease models has become increasingly important during the COVID-19 pandemic. The forecasting of disease spread using mathematical models has become a common practice by public health authorities, assisting in the creation of policies to combat the spread of COVID-19. Common approaches to the modeling of infectious diseases include compartmental differential equations and Cellular Automata, both of which do not easily model the spatial dynamics of disease spread over unique geographical regions. 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 guidelines, 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 citizens, but all precautions proved insufficient to prevent the pandemic. 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 serious pathogens such as Ebola and HIV (Rachah and Torres 2016; Djordjevic et al. 2018). One widely used method of modelling infectious disease spread is the use of 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 geographical regions with unique characteristics.

To investigate the spatial dynamics of disease spread, Cellular Automata (CA) models have been used to divide physical space into a cellular representation, with rules of interaction 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 use of regular neighborhood patterns in traditional 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 non-uniform relation between geographical areas, and even fewer have applied cellular geography in modelling of disease spread (Zhong

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

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 formal disease spread model which assessed the effect that variolation to smallpox could have on average life expectancy in France. (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. Every member of the population is assumed to have a constant rate of daily contacts with other people of any infection category which are sufficient to spread disease, following the 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 permanent immunity. A set of differential equations can be derived to describe how a population's daily interactions cause infection and immunity to disease over time:

$$\begin{cases} S = S(t) & s(t) = \frac{S(t)}{N} \\ I = I(t) & i(t) = \frac{I(t)}{N} \\ R = R(t) & r(t) = \frac{R(t)}{N} \end{cases} \tag{1}$$

$$\begin{cases} \frac{ds}{dt} = -\lambda s(t) i(t) \\ \frac{dr}{dt} = \gamma i(t) \\ \frac{di}{dt} = \lambda s(t) i(t) - \gamma i(t) \\ s(t) + i(t) + r(t) = 1 \end{cases} \tag{2}$$

The set of equations (1) define the percentage of the total population N in each of the infection states over time. The set of equations (2) define how Susceptible individuals become 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, regardless of infection category. These assumptions are not ideal because they do not account for individualistic human behavior and the variance in disease effects between individuals. The trajectory of a pandemic described by a SIR model is predictable: a population begins as nearly 100% Susceptible, the Infected rise and fall, and the number of Recovered increases over time as in the graph of Figure 1.

To increase the accuracy of the SIR model, other compartmental states have been used in research, including Exposed states (E) which describe the latency period preceding contagious infection, Quarantined

states (Q) which describe a period of reduced human contact while in a stage of infection, Asymptomatic states (A) which define individuals who are not aware of their infection but are contagious, and a Dead (D) state which models terminal illness and the permanent removal of individuals from the population. Such compartmental models are often designated by creating an acronym from the sequence of infection states modeled. Many combinations and variations of states have been used to model the spread of pathogens in both differential equation models and CA models.

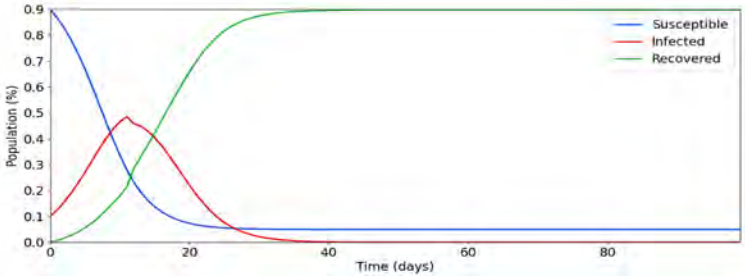


Figure 1: SIR Model Trajectory

The spread of COVID-19 using non-geographical SIR models has been successfully predicted 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) having two distinct phases of Infected was shown to predict the COVID-19 outbreaks in Wales and England. Many similar compartmental differential equation models have been shown to be accurate, but do not offer specific estimations of spatial disease spread dynamics.

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 stochastically generate heterogeneous cell spaces. The model includes geographical considerations in creating the cell spaces and uses a regular neighborhood pattern. A probabilistic SEIQR CA defined in (Ghosh and Bhattacharya 2020) use similar methods of cell space generation based on 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, based on the shape of their two dimensional geographical borders. (Cárdenas et al. 2020) introduced a SIR model built using the Cadmium Discrete Event System Specification simulation library, which is also used to implement the model presented in this paper.

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 of 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 an explicit delay function.

## 2.2 Geographical Cellular Models

The use of uniform neighborhood patterns in the cellular modelling of geographical phenomenon eliminates useful information about the unique features and relationships between regions. Allowing cells to have differing numbers of neighbors, sizes, and weights of influence facilitates the use of existing information about the geographical units under study. In a geography-based CA, the behavior of a cell's local transition function should be influenced by the shape, size, and characteristics of all 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 labeled polygons in the map of Figure 2, some variation can be seen in the distances between cell centroids and in the amount of border length shared between cells.

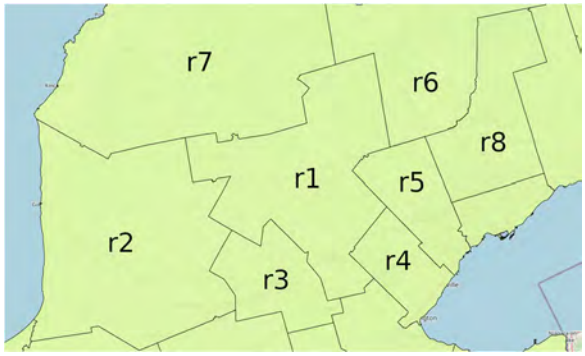


Figure 2: Geographical Cells

A simple method to define geographical neighborhoods is to consider all cells in the cell space as related by a correlational weight between every pair of regions, instead of a uniform neighborhood pattern. This geographical neighborhood pattern is greatly simplified by setting the correlational weight between regions that should not be related to zero. The method of geographical correlation we used in the model of this paper considers the amount of shared border length between regions, where regions that do not share borders have a correlational weight of zero. This geographical correlation can be calculated by dividing the length of border shared between two cells by the total border length of each region.

### 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 contagious. The length of the exposed state 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 that includes possible reinfection, the Recovered population has immunity from disease for a period before becoming Susceptible again. If reinfection is not modeled, then the Recovered state can be represented in a single phase. The Dead state describes the population that dies because of infection but do remain in the original population count.

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 correlational weights, $V_i = \{(m_i, 1), (m_k, c_{ik}),...,(m_x, c_{ix})\}$. 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 influence human behavior and disease outcomes. The probability that an Exposed individual enters the Infected state is controlled by the incubation rate $\varepsilon(q)$. The probability that an Infected individual becomes

Recovered is controlled by the recovery rate γ(p). The probability that an Infected individual enters the Dead state from each phase of Infected is controlled by the fatality rate δ(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 rate at which individuals make potentially infectious contacts with others is described by the mobility rate μ(p), which models the difference in movement behavior between age groups. The local transition function of the discrete time SEIRD model is defined by equations (3a)-(3g).

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

The capital A symbol in Equation (3a) is defined as the set of age groups within 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 newly Exposed population is equal to the product of the Susceptible population, 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 deaths 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 Dead equation (3f) states that the total deaths within a cell are equal to the deaths of the previous time step plus all new deaths occurring due to each phase of Infected. The Susceptible Equation (3g) 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 incubation rate $\varepsilon(q)$. The mean symptomatic infection length of COVID-19 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 recover in each phase of Infected (CDC 2021). The virulence and fatality rates of the model were varied to match infection case data in experimentation. The fatality rate of each age group in a cell depends on the cumulative sum of infections across all age groups of that cell, modelling the effect of an over capacity health care system. If the level of infection within a cell grows beyond a threshold, the fatality rate increases according to a fatality rate modifier. Several new factors 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. Contributing to $g_{ij}$ is the movement restriction factor $k_i$, which is introduced to define infection dependent mobility restriction on populations. Several movement restriction factors may be assigned to a single cell to define tiers of lockdown intensity, forming a set of $k_i$ values for a cell. When calculating the $g_{ij}$ between two cells, the more severe movement restriction factor in effect is used to describe their interaction. The geographical correlation and neighborhood definition remain the same as in section 2, and the rules of equations (3a)-(3g) still apply with some substitutions.

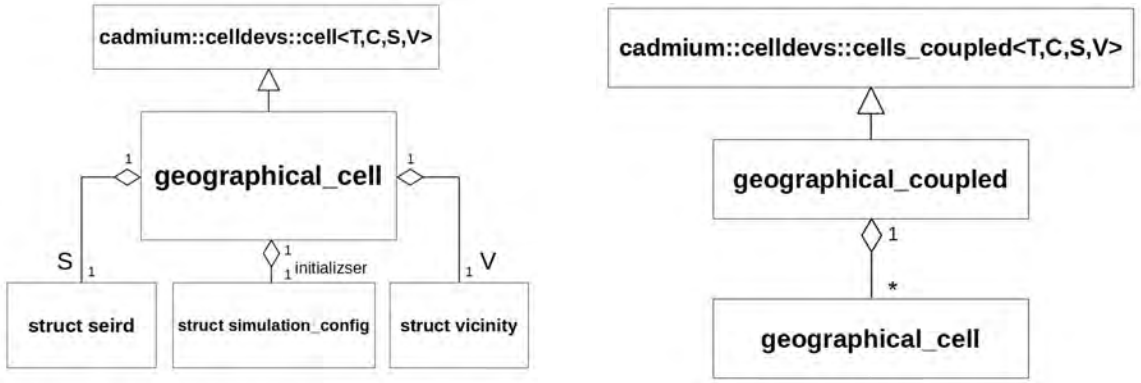


Figure 3: Geographical Cell model UML Class Diagrams

The infection threshold of a $k_i$ value defines the cumulative amount of active infection above which its mobility rate multiplier $\mu_k$ will be active. The mobility rate multiplier of each $k_i$ becomes inactive when the level of infection falls below the infection threshold level minus a hysteresis level. Only a single $k_i$ may be active at any time within a cell, 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 individuals in all age groups that ignore the movement restrictions in any $k_i$. When considering the interaction between two cells, the more restrictive infection correction factor is used.

The model is implemented as a Cadmium Cell-DEVS model called “geographical_coupled”, consisting of atomic cell models of type “geographical_cell” that are coupled to form a cell space. Each geographical_cell uses a cell id (C), a state object that contains all cell state variables (S), and a vicinity object that describes the structure of a cell’s neighborhood (V). The relationships between the classes involved in the Cell-DEVS model can be seen in the Unified Modeling Language (UML) diagram of Figure 3.

The infection state variables computed in the local transition function are discretized per equations (4a)-(4e), where P is the precision divider, and the use of square brackets denotes a round operation:

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

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 is returned in line 9 to be written as the new cell state.

Table 1: Geographical Cell local computation function

| **Input:** none. **Output:** New State (struct seird) |
|---|
| **1:** *new* = seird()<br>**2: for**(each age group *a* in *s.age_groups*)<br>**3:** new.E.at(a) = new_exposed(a,*s*)<br>**4:** new.I.at(a) = new_infected(a,s)<br>**5:** new.R.at(a) = new_recovered(a,s)<br>**6:** new.D.at(a) = new_dead(a,s)<br>**7:** new.S.at(a) = 1 - new.E.at(a) – new.I.at(a) - new.R.at(a) - new.D.at(a)<br>**8: endfor**<br>**9: return** new; |

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 for an age group b of a single cell i can be explained by the pseudocode in Table 2. The algorithm begins by calculating the movement correction multiplier of cell i, using the movement correction function which scans the set of movement restriction factors $k_i$ and chooses the most suitable factor to apply. The infection correction factor $g_i$ is calculated in line 2 by considering the proportion of disobedient d that will follow the movement restriction policy and the mobility restriction of the $k_i$ selected. In line 3 the variable infStrength is declared and is 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 contact made with all neighbors. This is done by first calculating the infection correction factor of the neighbor cell j, and then choosing the more restrictive correction factor between i and j to describe their interaction (lines 5-7).

Next the algorithm calculates the infective potential of each age group in neighbor cell j by summing over all infection phases 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.

Table 2: New Exposures Algorithm

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

## 4 EXPERIMENTAL RESULTS

The map of Figure 4 (OHESI 2021) shows the province of Ontario, Canada, including the 34 Public Health Units (PHUs) of Ontario that we used in testing the model defined in section 3.


Figure 4: Public Health Units 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 present at any time 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 10 days mean infection length of COVID-19 when converting case reports to active cases. 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 was provided by (Ontario GeoHub 2021).

The PHUs of Ontario were placed in 5 groups based on their location: Central East, Central West, Southwest, East, and North. A visualization of the case data over for these groups is shown in Figure 5, where the North PHU group is omitted because of space constraints. The default experimentation parameters used in experimentation as well as the specific structure of the cell state variables can be found at https://github.com/SimulationEverywhere.

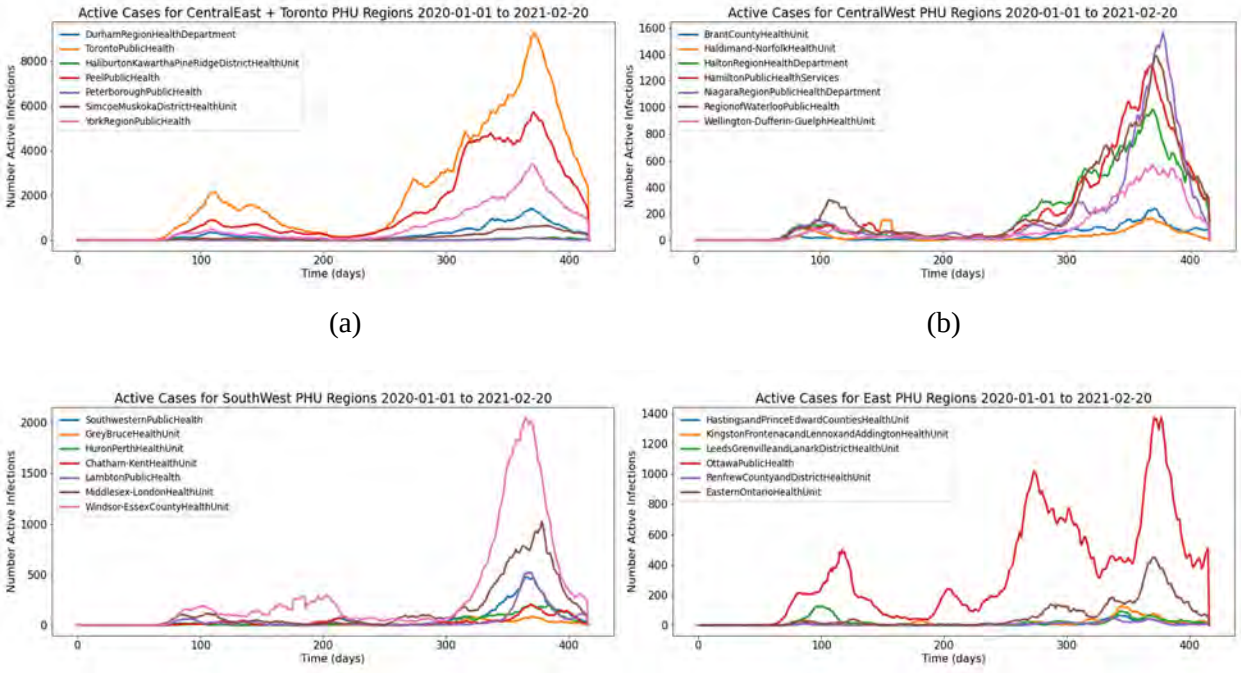


Figure 5: Ontario PHU Case Data from Dec 2020 to February 2021: (a) Central East PHUs; (b) Central West PHUs; (c) Southwest PHUs; (d) East PHUs.

We examined the effect of movement restriction policies in the Toronto PHU in terms of the movement correction factor's infection thresholds, hysteresis values, and mobility multipliers. Without movement restriction policies to limit infection spread, most of the population in all 34 cells become rapidly infected in a severe trajectory.

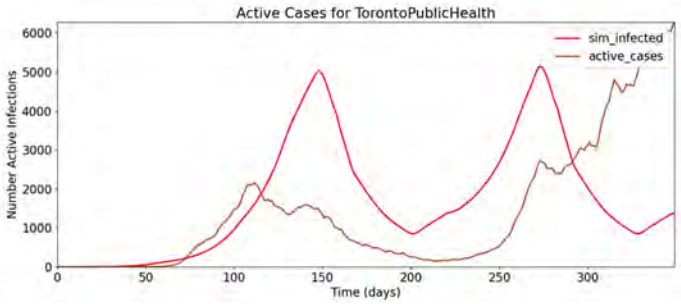


Figure 6: Toronto Pandemic, Movement Restriction Factor #1

Figure 6 shows experimental simulation data compared to the reported active cases. We used an infection threshold too high in value, resulting in movement restrictions that are applied later than what is

represented in the case data. The predicted active infections would have been more accurate had the infection threshold of the limiting $k_i$ policy been in effect at half the number of active infections.

By using an infection threshold 50% lower than the simulation of Figure 7, the movement restriction policy comes into effect at a more accurate level of infection, as shown in Figure 7.

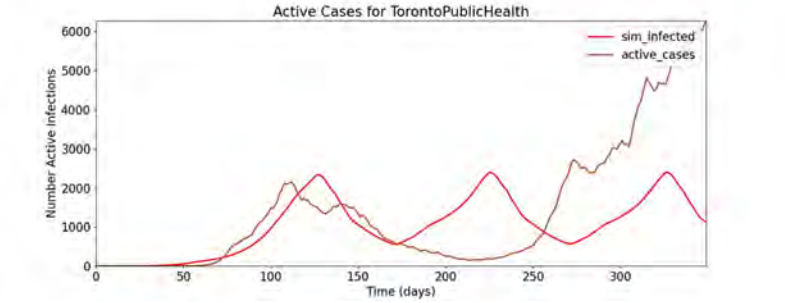


Figure 7: Toronto Pandemic, Movement Restriction Factor #2

However, these restrictions are lifted too soon at day 170 of the simulation, 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 dominant infection correction factor, 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 8.

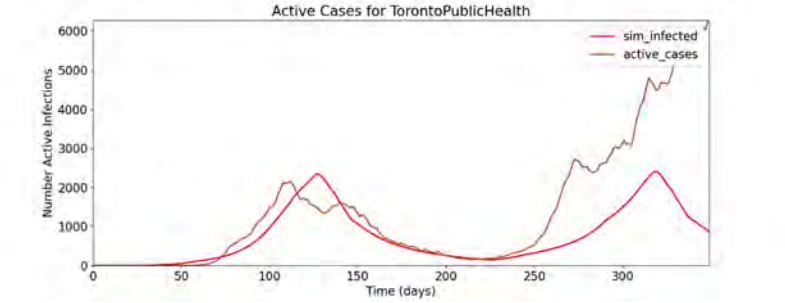


Figure 8: Toronto Pandemic, Movement Restriction Factor #3

After these preliminary tests, we conducted a series of simulations and compared the model results 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 the virulence rate above the default level.

While the first wave of the COVID-19 pandemic in Ontario 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 which also had considerable levels of first wave infection. Figure 9 shows the results in the regions surrounding Toronto, Ottawa, and Windsor Essex.

The simulations in these regions do track the general shape of the case data, but in some results there is a 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. Every cell is using the same set of infection correction factors, which could be a significant source of error because each region's population density and human characteristics are region specific by nature. In fact, in the case of regions directly adjacent to Toronto, the level of simulation error is proportional to the population density

of each region. 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. Given that Toronto has both the highest population density and highest model accuracy, and that Durham has both the lowest population density and accuracy, a population density correction should be investigated in the future.

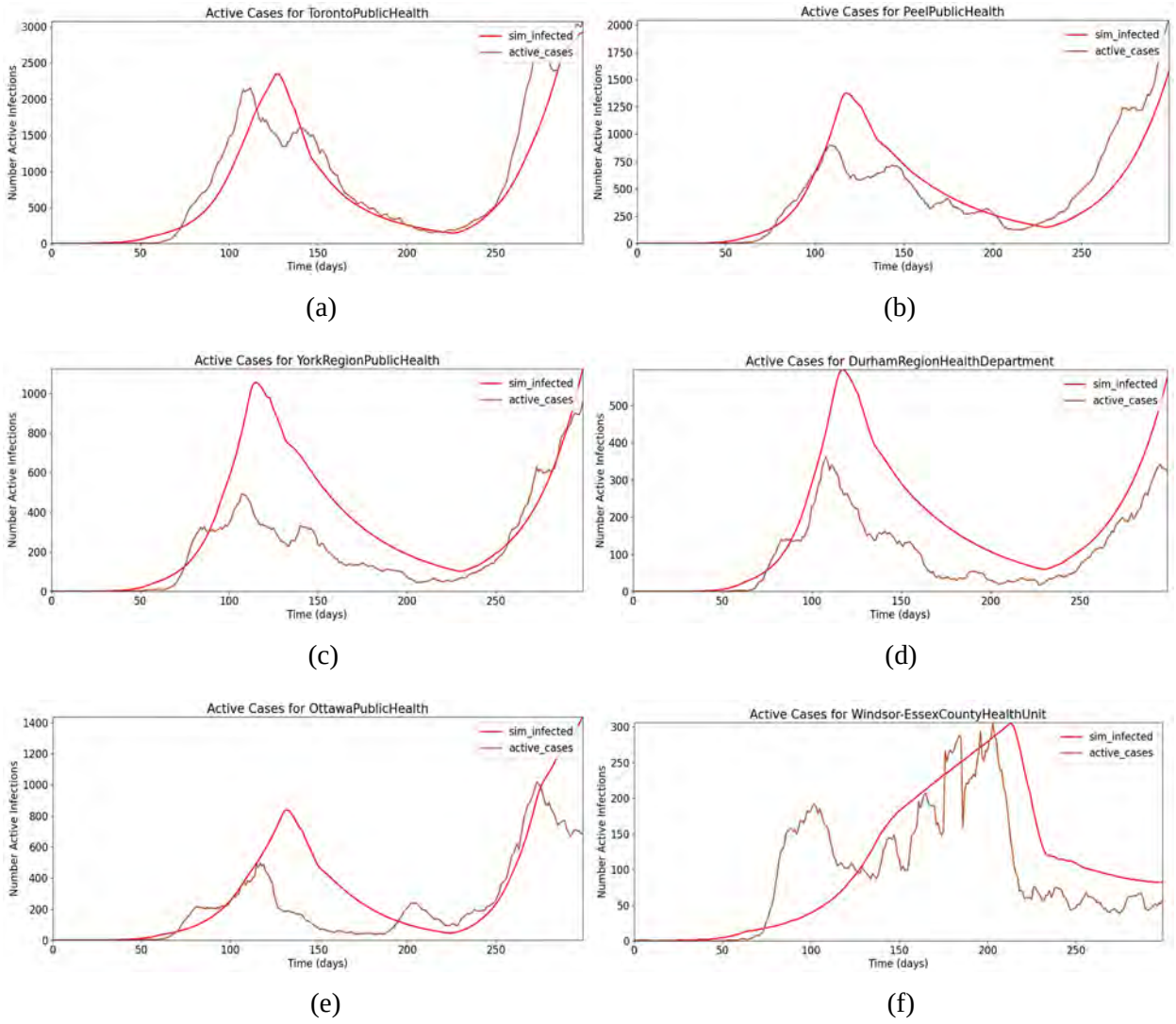


Figure 9: Pandemic in (a) Toronto; (b) Peel; (c) York; (d) Durham; (e) Ottawa; (f) Windsor Essex.

## 5 CONCLUSIONS

A Cell-DEVS COVID-19 model was defined and simulated to study the spread of infectious disease in geographical environments. The model defined in this paper uses a method of geographical CA like that used by (Zhong et al. 2009) in their simulation of the original SARS pandemic, but considers more factors such as movement restriction policies. The model was simulated using the geography of Ontario, Canada, by dividing 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 region specific disease characteristics and the addition of a population density correction factor.

## REFERENCES


Caccavo, D. 2020. *Chinese and Italian COVID-19 Outbreaks can be Correctly Described by a Modified SIRD Model. MedRxiv*. https://www.medrxiv.org/content/10.1101/2020.03.19.20039388v1.full.pdf, accessed 2nd February 2021.

Cárdenas, R., K. Henares, C. Ruiz-Martín, and G. Wainer. 2021. "Cell-DEVS Models for the Spread of COVID-19". In *Proceedings of the 2020 International Conference on Cellular Automata for Research and Industry*, December 2nd–4th, Łódź, Poland, 239-249.

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*. https://www.medrxiv.org/content/10.1101/2020.02.12.20022566v1, accessed 2nd February 2021.

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". In *Proceedings of the Royal Society of London A*, August 1st, London, England, 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, 6-15.

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