Geographical SEIRDS COVID-19 Model

Glenn Davidson
Department of Systems and
Computer Engineering
Carleton University
Ottawa, Ontario, Canada
glenn.davidson@carleton.ca

ABSTRACT

An existing Geographical SIRDS Cell-DEVS model was adapted to include an Exposed stage that follows the statistical distribution of virus incubation time in COVID-19. Cell-DEVS is a discrete-event Cellular Automata formalism, based on the Discrete-Event System Specification and Cellular Automata theory. The Cadmium JSON library is used to define the model and initialize the cell space with complex state variables and neighborhoods based on real geographical areas. This paper describes methodologies of pandemic modeling, geographical modeling, and demonstrates how Cell-DEVS and the Cadmium JSON library can be used to develop geographical cellular models. The parameters of the model have yet to be adjusted to match real pandemics, but the results of completed simulations are intuitive. The framework to input a realistic infection profile is complete, at which time the model can be compared to and calibrated by existing pandemic data in the public domain.

1 Introduction

Beginning in late 2019, a new coronavirus pandemic emerged that has presently infected the global population, resulting in significant economic impacts and 1.8 million deaths worldwide [1]. The new virus officially named SARS-CoV-2, otherwise known as COVID-19, is just one of the many diseases that the human race has been in conflict with for the entirety of our species evolution. Bacteria, viruses, fungi, and parasites are all infectious agents, capable of infecting a large number of people and causing mass fatalities. Never in the history of our species have these pathogens been able to travel the entire globe at rapid pace. The COVID-19 pandemic highlights the importance of studying and combatting the spread of pathogens, so that their effect on daily life may become negligible. If the mechanisms of disease spread are understood, efficient interventions may be implemented to contain local pandemics before they spread globally, and to contain transmission in the event of a global pandemic. The study of pandemics and pathogens can be used to develop a predictive model of how particular infectious agents would spread in specific geographical environments over time, which can assist Governments and scientists in planning and implementing containment policies.

An existing Cell-DEVS geographical COVID-19 pandemic model [2] has been extended and is presented in this paper. The existing model is based upon a discrete time SIRS Cellular Automata model [3]. Cell-DEVS uses the theory of Cellular Automata (CA) and the Discrete-Event Simulation Formalism (DEVS), to form a toolkit for the modeling and simulation of cell-based models. The extended model presented is geographical Susceptible, Exposed, Infected, Recovered, Fatal (SEIRDS) model with optional reinfection capability. The SEIRDS model also utilizes the Cadmium JSON library to define the model, and is used to load pre-defined neighborhoods and model data at runtime. The Cadmium JSON library is advantageous because it allows complex cellular data to be loaded from input files at run time, resulting in a flexible model that does not need to be edited when running a large variety of scenarios.

The model discussed in this paper considers the population of each geographical cell as divided into age groups, each of which may have different infection characteristics and behaviors. All members of each age group are in one of the states S, E, I, R, D at any time. The model considers infection incubation over time, virulence rate over time, and recovery rate over time. Also considered are the mobility rates of per age group, fatality rates per age group, the effect of lockdown interventions, and the increase in fatalities associated with an over capacity health care system.

2 Background

2.1 Cellular Automata Pandemic Models

Pandemics were first modeled as a set of differential equations in 1927, dividing a population of interest into one of three categories: Susceptible (S), Infected(I), and Recovered(R) [4]. Susceptible people are vulnerable to the pathogen, and will become infected if they have a sufficient contact with infected people. Infected people have the disease under investigation, and spread the disease to Susceptible people through their daily contacts with the population. Infected people have a certain probability of becoming Recovered each day, defined as the recovery rate. Recovered people have overcome the disease, are no longer infectious, and are permanently immune to the disease. The typical trajectory of a SIR model pandemic is shown in Figure #1, where the curves are expressed as percentages of the total population in each state.

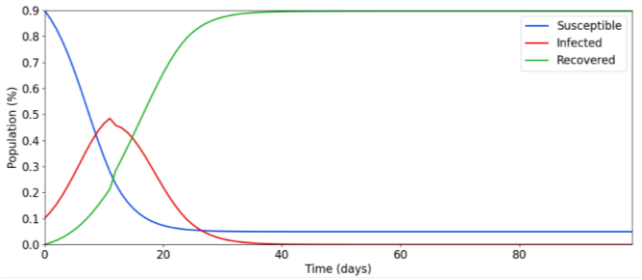

Figure 1: SIR Model Trajectory

With these categories in place, a set of differential equations can be derived to describe the infection state of the population over time. The S, I, and R states can be expressed as a function of time, and in percentages of total population instead of in concrete numbers:

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

If some assumptions are made about the behavior of the population, then the set of differential equations (2) become quite simple.

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

This model assumes no death or birth in the population, and assumes every person behaves identically, having a constant number of daily contacts at random, regardless of any person's infection status. These assumptions make such a simple model impractical for predicting the real world. To increase the accuracy of this model, more infection states can be added, such as a Fatality, Exposed, Passive Immunity, Vaccinated, and the possibility of reinfection after recovery. The infection states that do vary over time in the real world (Exposed, Infected, Recovered) can be modeled with sub-states that represent sequential days of infection. Real world data can then be used to profile how the infection behaves on average per infection day, from initial exposure to eventual recovery or fatality.

Populations in the Exposed state can transition to the Infected state according to a statistical curve of the pathogen's incubation time. Each day of the Infected state can have variable contagiousness, probability of recovery, and probability of fatality. A state transition diagram of a SERIDS model is given in Figure #2. Adding a second S to the end of the model type name indicates that the Recovered population becomes Susceptible again after a number of days spent in the Recovered state.

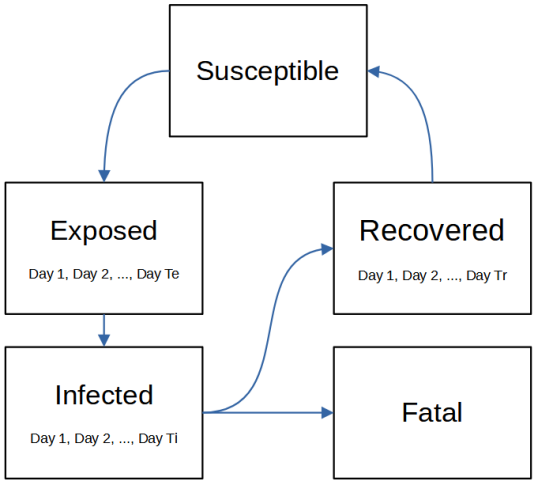

Figure 2: SEIRDS State Diagram

A pathogen's incubation period is defined as the number of days before an infected person shows symptoms, and is an important variable in predicting disease. A meta-analysis of the incubation period of COVID-19 was used to generate a symptom onset profile representing the probability of becoming Infected on each day. The incubation period in COVID-19 has been shown to follow a log-normal distribution, with a mean onset time of 5.1 days and showing as late as 24 days, shown in Figure #3 [5].

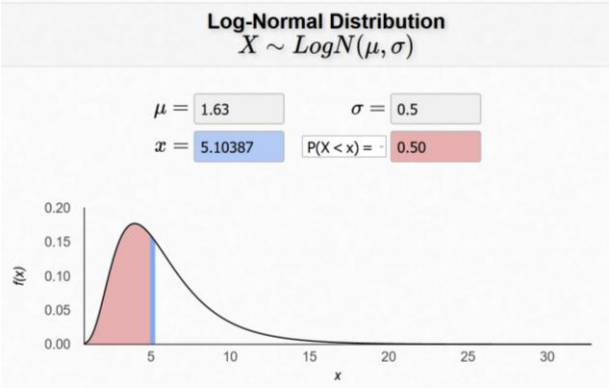

Figure 3: COVID-19 Incubation Time Distribution

2.2 Cellular Automata Pandemic Models

The models discussed so far had no consideration of physical space and therefore geographical spread of pathogen. Cellular Automata (CA) can be used together with differential equation models to create a spatial pandemic model. CA is a mathematical formalism that describes an N dimensional space of adjacent "cells" that contain values that can change over time. Any number of dimensions may be represented in CA, but the most practical models are in either two or three dimensions. The value of each cell is computed by the same set of transition rules executed in discrete time steps, which describe how the value of neighboring cells affect each other.

SIR based discrete time CA models have widely studied [3]. However traditional CA may not provide realistic neighborhoods when modeling large and complex geographical areas. Using real geographical data to define cell spaces is preferred to hypothetical scenarios because the model results are verifiable to a degree, and offer more useful intuition about the world. However, the question of how to best model geographical relation is difficult. A simple heuristic principle of geographical relation is: Every geographical area is related to every other geographical area, but near areas are more related than distant areas [6]. Using this principle, a correlation factor can be calculated to represent the level of relation or connection between a pair of geographical areas.

The simplest possible correlation factor would consist of a binary 0 or 1, indicating no correlation and correlation respectively. Ultimately the method of geographical relation depends on the subject of study. SIR based models track infection of a geographically dispersed population, so the correlation factor should represent the flow of population between geographical areas.

A correlation factor for population flow between two regions could use the length of border shared between the regions, and divide the shared length by the total border length of each region. Transportation infrastructure such as roads, highways, and airports could also be considered and combined with this approach. A correlation factor c_{ij} based on shared border between two cells i and j can be calculated as:

$$\begin{cases} w_{ij} = \frac{z_{ij}}{l_i} & (3a) \\ w_{ij} = w_{ji} = \frac{\frac{z_{ij}}{l_i} + \frac{z_{ji}}{l_j}}{2} & (3b) \\ c_{ij} = w_{ij} & (3c) \end{cases}$$

Equation (3a) describes a weight w_{ij} between two cells i and j based on shared border length. This weight is not identical in both directions of correlation between the two cells. Equation (3b) defines a weighting factor for cells i and j that are identical in both directions of correlation. Equation (3c) states that the correlation factor is equal to the weighting factor in this case, and no additional computation is required to scale w_{ij} and w_{ji} to the range of 0 to 1.

The correlation factor of (3c) can be used to describe how strongly a Susceptible population of one cell interacts with the Infected population of another cell. Cell-DEVS neighborhoods can be defined by (3c) instead of by predetermined neighborhood patterns, where a cell is in another cell's neighborhood if there exists a correlation factor between them that is greater than 0. A neighborhood defined this way allows cells to have any number of neighbors.

The SIRDS model on which the model of this paper is originally based also uses this method of neighborhood construction and shared border length correlation factors. The paper of [2] describes a discrete time CA model developed to investigate the geographical spread of SARS-COV-1 in China. The SEIRDS model described in this paper is similar in the infection states used, but adds more details to the infection profile and also uses different simulation methodologies and tools.

This model of this paper uses Cell-DEVS, and the Cadmium JSON library to implement and simulate the model. Cell-DEVS is a discrete-event Cellular Automata formalism, which utilizes the Discrete Event System Specification [7] (DEVS) and theory of Cellular Automata [8].

The Cadmium JSON library enables the user to load cells from a JSON format file, allowing the user to generate complex neighborhoods and state variables to be run as scenarios of the model. This project generates a Cadmium JSON from geographical data and state information before using it as an input to the model. The 2016 Canadian census offers good data on populations per geographical regions and is used in generating the results of this model. The input JSON is generated by python script, which parses geographical census data from input files and creates a cell space with neighborhoods corresponding to the geographical data, virus characteristics, and initial pandemic conditions. For each geographical area in the input data, a cell id is created, a set of cell state variables are initialized, and the neighborhoods of each cell are constructed by correlation factor.

3 Geographical SEIRDS Specification

The Geographical SEIRDS model can be described as a coupled Cell-DEVS model, where the cell space is populated by geographical cells of the same structure, each representing a unique geographical area. A geographical cell consists of a cell ID, a set of SEIRDS model state variables, a neighborhood, and simulation configuration data as depicted in the class diagram of Figure #4. Each neighbor relation pair uses a unique vicinity struct that describes the relationship in terms of cell IDs, geographical correlation factor, and infection correction factors. All geographical cells are placed in a top level coupled cell model called geographical_coupled, shown in the class diagram of Figure #5. At model runtime, the geographical_coupled model is initialized with the cell data in the input JSON file by using methods of the top model class cadmium::celldevs::cells_coupled<T,C,S,V>.

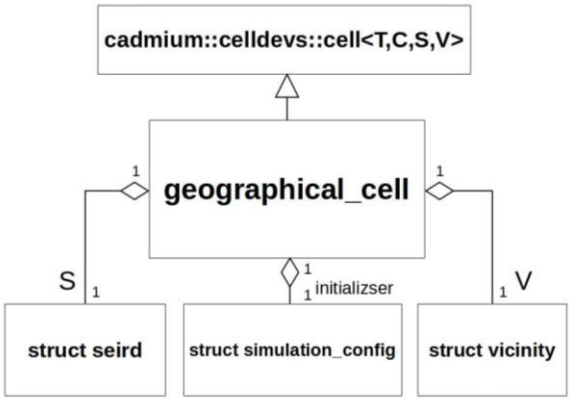

Figure 4: Geographical Cell UML Class Diagram

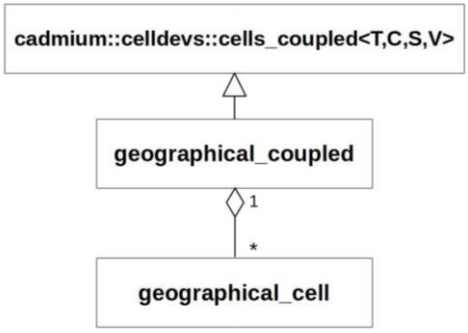

Figure 5: Geographical Coupled UML Class Diagram

The variables in the simulation_config struct (Figure #6) describe the pathogen under investigation, so the same configuration struct is used to initialize all cells in the cell space. The variable prec_divider defines the level of precision used in the simulation, and is referenced as the A value in equation (4).

```
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;

    bool SIIRS_model = false;
};
```

Figure 6: simulation_config struct

The seird struct (Figure #7) serves as the state object of class geographical_cell, and its attributes are the subject of the transition rules given in equations (5) to (9).

```
struct seird {
    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>> recovered;
    std::vector<double> fatalities;
    std::unordered_map<std::string, hysteresis_factor> hysteresis_factors;

    double population;
    double disobedient;
    double hospital_capacity;
    double fatality_modifier;
```

Figure 7: struct serid State Variables

The top coupled model geographical_coupled is composed of a set of k geographical regions K = {1,2,...k}, organized in arbitrary order. These unique geographical regions are represented as a cell space of geographical cells M = {m_1, m_2, ..., m_k}. Each cell in M is considered to have a population consisting of an integer number of individuals that are capable of contracting and spreading COVID-19. The members of the population are divided into different states of infection: S, E, I, R, D. These state variables describe the percentage of the total cell population that are in that state. Each cell's population is divided into a number of age segments so that age specific behaviour may be defined. We can describe the infection state of age group a in cell i at time t by the following notation: $S_{i,a}^t$, $E_{i,a}^t$, $I_{i,a}^t$, $R_{i,a}^t$, and $F_{i,a}^t$.

The Cell-DEVS cell delay used for simulation is invariantly a transport delay of one day, meaning that all cells get inputs from neighbors on the transition of 1 day. The discussion of infection over time will herein be referenced using an effective time step of 1 day. The states Exposed, Infected, and Recovered describe time varying behaviour and are implemented as a series of sub-states, each representing a sequential day within the infection state. The Susceptible and Fatal state lengths are indefinite, and are described without sub-states.

The length of each state is fixed and identical for all cells. T_e, T_i, and T_r represent the length in days of the states Exposed, Infected, and Recovered. The state set of the infection a cell population undergoes can be represented as a set of states P = {0, 1, ..., T_e + T_i + T_R + 1}, where state 0 is Susceptible, state 1 is the first day of Exposed, state T_e + 1 is the first day of Infected, state T_e + T_i + 1 is the first day of Recovered, and state T_e + T_i + T_r + 1 is Fatal. To distinguish which sub-state is referred to in the transition rules, the variable q is used to denote the infection sub stage when referring to states that consist of multiple days. For example $E_{i,a}^t(q)$ is the proportion of age group a in cell i with Exposed state q at time t, where q $\in$ {1, 2, ..., T_e}.

To describe total state set Q of the model, the state variables must be discretized so that Q is finite. The discretization of the state variables S, E, I, R, D, per age group can be described by the following set of equations:

$$\begin{cases} DS_{i,a}^{t} = \dfrac{\left[AS_{i,a}^{t}\right]}{A} & (4a) \\ DE_{i,a}^{t}(q) = \dfrac{\left[AE_{i,a}^{t}(q)\right]}{A} & (4b) \\ \qquad q \in [1, 3 \ldots, \mathrm{T}_e] & \\ DI_{i,a}^{t}(q) = \dfrac{\left[AI_{i,a}^{t}(q)\right]}{A} & (4c) \\ \qquad q \in [\mathrm{T}_e + 1, \mathrm{T}_e + 2, \ldots, \mathrm{T}_e + \mathrm{T}_i] & \\ DR_{i,a}^{t}(q) = \dfrac{\left[AR_{i,a}^{t}(q)\right]}{A} & (4d) \\ \qquad q \in [\mathrm{T}_e + \mathrm{T}_i + 1, \mathrm{T}_e + \mathrm{T}_i + 2, \ldots, \mathrm{T}_e + \mathrm{T}_i + \mathrm{T}_r] & \\ DF_{i,a}^{t} = \dfrac{\left[AF_{i,a}^{t}\right]}{A} & (4e) \end{cases}$$

Where A denotes the precision used and the use of square brackets indicates a round operation. The set of all possible combinations of DS, DE, DI, DR, and DE form the finite state set Q.

The transition rules for each cell determine its next infection state. The transition rules can be divided into two categories, internal effects which depend only on a cell's internal state, and external effects that depend on both the internal state of a cell and the state of all neighbors. There are two sets of transition rules for this model, one that enables reinfection, and one that disables reinfection. If reinfection is enabled, then the Recovered population becomes Susceptible again after finishing the recovered state. The transition rules for the SEIRD and SEIRDS models are given by the following set of equations and symbol table:

$$\begin{cases} F_{i,a}^{t+1} = F_{i,a}^{t} + \sum_{q \in \{T_e+1, T_e+2, \ldots, T_e+T_i\}} \mathfrak{f}_a(q)\, I_{i,a}^{t}(q) & (5a) \\ E_{i,a}^{t+1}(1) = \sum_{\substack{j \in \{1,2,\ldots,k\} \\ q \in \{T_e+1, T_e+2, \ldots, T_e+T_i\}}} c_{ij} k_{ij} \lambda_a(q) \mu_a(q) S_{i,a}^{t} I_j^{t} & (5b) \\ E_{i,a}^{t+1}(q) = \big(1 - \varepsilon_a(q-1)\big) E_{i,a}^{t}(q-1) & (5c) \\ \qquad q \in \{2, 3, \ldots, T_e\} & \\ I_{i,a}^{t+1}(T_e+1) = E_{i,a}^{t}(T_e) + \sum_{q \in \{1,2,\ldots T_e - 1\}} \varepsilon_a(q)\, E_{i,a}^{t}(q) & (5d) \\ I_{i,a}^{t+1}(q) = I_{i,a}^{t}(q-1)(1 - \gamma_a(q-1) - \mathfrak{f}_a(q-1)) & (5e) \\ \qquad q \in \{T_e + 2, T_e + 3, \ldots, T_e + T_i\} & \\ R_{i,a}^{t+1}(T_e + T_i + 1) = I_{i,a}^{t}(T_e + T_i) & (5f) \\ \qquad + \sum_{q \in \{T_e+1, T_e+2, \ldots, T_e+T_i-1\}} \gamma_a(q)\, I_{i,a}^{t}(q) & \end{cases}$$

$$\begin{cases} R_{i,a}^{t+1}(q) = R_{i,a}^{t}(q-1) & (5g) \\ \qquad q \in \{\mathrm{T}_e + \mathrm{T}_i + 2, \mathrm{T}_e + \mathrm{T}_i + 3, \ldots, \mathrm{T}_e + \mathrm{T}_i + \mathrm{T}_r\} & \end{cases}$$

$$\begin{cases} R_{i,a}^{t+1}(q) = R_{i,a}^{t}(q-1) & (5h) \\ \qquad q \in \{\mathrm{T}_e + \mathrm{T}_i + 2, \mathrm{T}_e + \mathrm{T}_i + 3, \ldots, \mathrm{T}_e + \mathrm{T}_i + \mathrm{T}_r - 1\} & \\ R_{i,a}^{t+1}(\mathrm{T}_e + \mathrm{T}_i + \mathrm{T}_r) = R_{i,a}^{t}(\mathrm{T}_e + \mathrm{T}_i + \mathrm{T}_r) & (5i) \\ \qquad + R_{i,a}^{t}(\mathrm{T}_e + \mathrm{T}_i + \mathrm{T}_r - 1) & \end{cases}$$

$$\begin{cases} S_{i,a}^{t+1} = 1 - \sum_{q=1}^{T_e} E_{i,a}^{t+1}(q) - \sum_{q=T_e+1}^{T_e+T_i} I_{i,a}^{t+1}(q) & (5j) \\ \qquad - \sum_{q=T_e+T_i+1}^{T_e+T_i+T_r} R_{i,a}^{t+1}(q) - F_{i,a}^{t+1} & \end{cases}$$

Symbol:	Definition:
$\mathfrak{f}_a(q)$	Fatality rate of Infected stage q for age group a, after correcting for hospital capacity effects
$\lambda_a(q)$	Virulence rate for Infected stage q of age group a
$\mu_a(q)$	Mobility rate during Infected stage q for age group a
c_{ij}	Geographical Correlation factor between cell i and j
k_{ij}	Correction factor applied to both cell i and j
$\varepsilon_a(q)$	Incubation rate of exposed stage q for age group a
$\gamma_a(\mathrm{q})$	Recovery rate of Infected stage q for age group a

Table 1: Transition Rules Symbols

Equation (5g) is used for calculating the Recovered if re-infection is enabled, and Equations (5h) and (5i) are used in place of (5g) if re-infection is not enabled. A description of the transition rules from the perspective of age group a in cell i is:

Equation (5a) states that the Fatal of the next day is equal to the Fatal of the current day plus the sum of all Infected that die when moving to the day. The summation in (5a) states that the newly deceased population is equal to the sum of the infected population of each infection day multiplied by the fatality rate of that infection day.

Equation (5b) describes the next day's new Exposed as a result of contacts between the Susceptible population of cell i and Infected population of all neighboring cells. The summation depicts the Susceptible population of age group a in cell i coming into contact with the total Infected population (all age groups) of each of its j neighbors. Note that cell i includes itself in its own neighborhood, so that the Susceptible of cell i do interact with the Infected of cell i internally. The virulence rate $\lambda_a(q)$ represents how infectious the Infected population is on stage q of Infected. Cell i is related to each of its neighbors through their geographical correlation factor c_{ij}, which describes the amount of interaction that the pair of cells exchange. The correction factor k_{ij} alters the correlation between cells if any lockdown policies are in effect.

Equation (5c) describes how the Exposed population advances towards the Infected state. The Exposed in stage q is equal to the exposed of the previous day multiplied by $1 - \varepsilon_a(q-1)$, where $\varepsilon_a(\mathrm{q})$ describes the probability of leaving state q of Exposed and entering Infected day 1.

Equation (5d) describes the new Infected entering Infected day 1 from all Exposed stages. On any day of Exposed, a proportion will enter Infected on the next day according to the incubation rate $\varepsilon_a(q)$. The population on the final day of Exposed all enter day 1 of Infected on the next day.

Equation (5e) states that the Infected on all days other than Infected day 1 is equal to the Infected of the previous day, minus those that become Recovered or Fatal.

Equation (5f) states that the new Recovered are equal to the proportion of Infected at each stage that become recovered according to the recovery rate $\gamma_a(q)$, plus the Infected that reached the last infection stages.

Equation (5g) describes progression through the Recovered stages if re-infection is enabled. It states that for all Recovery stages except the first, the Recovered of the present day is equal to the recovered of the previous day.

Equation (5h) and (5i) describe the progression through the Recovered stages if re-infection is disabled. In this case the Recovered will advance to the last Recovered stage and stay there indefinitely.

Equation (5j) uses the fact that the sum of all states S, E, I, R, D is equal to one to calculate the Susceptible of the next day. Note that the E, I, R, and D of the next day are calculated before using this formula.

The model also defines an infection threshold at which the fatality rate of infection will increase, defined as the Hospital Capacity. This number is defined as a percentage of the population size. If the percentage of Infected population exceeds the hospital capacity in a cell, then all fatality rates in that cell are multiplied by a constant fatality rate multiplier that increases the fatality rate. The net fatality rate is calculated by:

$$\mathrm{f}_a(q) = \begin{cases} f_a(q) & if\ \sum I < Hospital\ Capacity \\ f_a(q)F & if\ \sum I \geq Hospital\ Capacity \end{cases} \quad (6)$$

Where $f_a(q)$ is the base fatality rate for an age group and F is the constant fatality rate multiplier.

The correlation correction factor k_{ij} reduces the correlation between cells if either cell has a lockdown policy in effect. The lockdown policy of the current cell i and neighbor cell j are first evaluated before choosing the more restrictive of the two as k_{ij}. The correction factor m_c of a cell is calculated by considering the current status of Infection within the cell, and the percentage of the population that follow lockdown policy. The disobedient proportion (d) of a cell's population is unaffected by lockdown policy, while the remainder of the population does follow policy. The correlation correction factor k_{ij} is calculated by using equations (7), (8), and (9).

$$k_{cell} = \mathrm{d} + (1 - \mathrm{d}) * m_c \quad (7)$$

$$k_{ij} = \min(k_i, k_j) \quad (8)$$

Lockdown policy is determined by the correction factors of a cell. A correction factor consists of an Infection Threshold I_{TH}, mobility correction factor c_m, and a Hysteresis level H. Correction factors are stored in a cell's vicinity struct seen in Figure #8. The form of a correction factor in the JSON format is $m_c = "I_{TH}": [c_m, H]$. A cell may have any number of correction factors that describe different behaviour at different levels of Infected.

$$m_c = \begin{cases} 1 & if\ no\ Lockdown \\ c_m & if Lockdown \end{cases} \quad (9)$$

If no lockdown policy is defined for a cell, then a default correction factor of "1": [1, 0] is used. Total infections can never exceed 100% of the population so this factor never is put into effect. Note that the geographical correlation factor (double correlation) is also stored within a vicinity struct.

```
struct vicinity
{
    using infection_threshold = float;
    using mobility_correction_factor = std::array<float, 2>;
    // The first value is the mobility correction factor mc;
    // The second one is the hysteresis factor H.

    std::map<infection_threshold, mobility_correction_factor> correction_factors;

    double correlation = 1.0f;
    explicit vicinity(double correlation) : correlation{correlation} {}
    vicinity(){}
};
```

Figure 8: struct simulation_config

4 Results

The SEIRD(S) Model was first simulated in a single geographical cell and then compared to the SIRD(S) model that the Exposed state was added to. The model was then simulated using the 2016 Canadian Census Dissemination Area (DA) data on the population and geography of Ottawa Ontario. Due to the Exposed state being added recently, the GIS Simulation Explorer [9] is currently unable to visualize the model's set of state variables in a map upon which the data is based, so results are given in the form of SEIRD(S) graphs in this paper. The Fatality graphs have been omitted in simulation comparisons where they are very similar.

4.1 Single Cell SIRD and SEIRD Simulations

The single celled simulations test the effect of lockdown in a SIRD model and a SEIRD model. The SIRD model simulations begin with 10% of it's population as Infected and 90% of its population as Susceptible. The SEIRD model simulations begin with 10% of it's population as Exposed, and 90% of its population as Susceptible. The behavior of the two models without lockdown are seen to be very similar seen in Figures #9-10. The Exposed state only delays the infection and reduces the peak Infected, while the

total recovered at steady state is visually identical. The model parameters for the single celled simulations are:

Population Size: **N**	1000
Age Groups: $\mathbf{N_A}$	5
Exposed State Length: $\mathbf{T_e}$	5
Infected State Length: $\mathbf{T_i}$	12
Recovered State Length: $\mathbf{T_r}$	16
Virulence Rates: $\boldsymbol{\lambda_a(q)}$	0.05 for T_i days all ages
Incubation Rates: $\boldsymbol{\varepsilon_a(q)}$	0.2 for T_e days all ages
Recovery Rates: $\boldsymbol{\gamma_a(q)}$	0.07 for T_r days all ages
Mobility Rates: $\mu_a(q)$	0.6 for T_i days all ages
Fatality Rates: $\mathfrak{f}_a(q)$	0.005 for T_i days all ages
Disobedient: **d**	0%
Hospital Capacity:	0.2
Fatality Modifier: **F**	1.5
Reinfection: **isSIRDS**	false
Infection Correction Factor: $\boldsymbol{k_{ij}}$	varies per simulation

Table 2: Single Celled Simulation Parameters

Table 2 defines what will be referred to as the base simulation parameters for the later SEIRD simulations, with the addition of a correction factor of "0.2": [0, 0.1] in simulations to do implement lockdown policy.

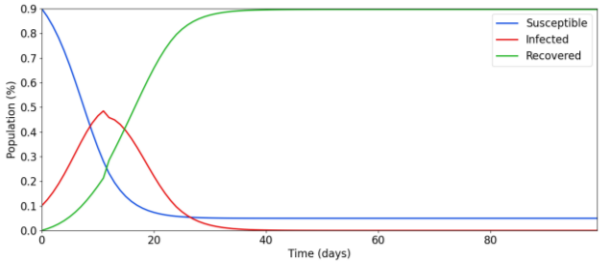

Figure 9: Single Cell SIRD No Lockdown

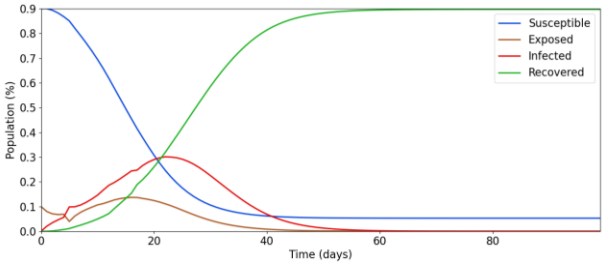

Figure 10: Single Cell SEIRD No Lockdown

Next a lockdown policy is introduced to both of the single celled simulations, using an infection correction factor of "0.2": [0, 0.1]. This corresponds to a lockdown threshold of 20% infected population, mobility multiplier of 0, and hysteresis of 0.1. This means that the mobility of each person becomes 0, no new infections will occur in the SIRD model, and no new exposures will occur in the SEIRD model. The lockdown policy will be removed when the total infected drops below 0.2 minus the hysteresis factor; at 0.1 infected. The results of these simulations are:

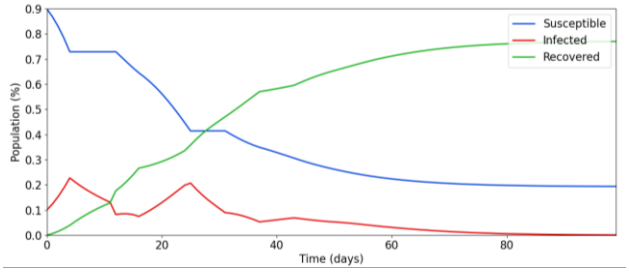

Figure 11: Single Cell SIRD With Lockdown

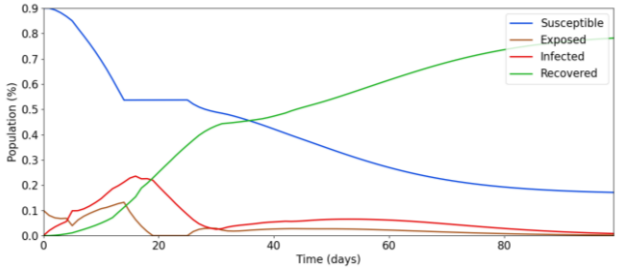

Figure 12: Single Cell SEIRD With Lockdown

The SIRD model is seen to enter lockdown after day 5, and undergoes two lockdowns in total during the simulation run. The SEIRD model only encounters one lockdown when initializing the same number of infections as starting in Exposed. The SIRD model's Infected population only briefly exceeds 0.2 as this lockdown assumes perfect isolation. When the lockdown engages in the SEIRD model, infections still increase for 5 more days as the Exposed population becomes Infected. After 5 days of lockdown, the Exposed population falls to 0 as there are only 5 Exposed phases.

The SIRD simulation reaches steady state at approximately 20 days before the SEIRD simulation.

4.2 Geographical SEIRD Simulations

Multi-celled Geographical scenarios are considered for the SEIRD model using the dissemination areas of Ottawa, Ontario. The base SEIRD parameters are used in the first multi-celled scenario, and a log-normal Exposed state is used with the base SEIRD parameters for the second scenario. The Exposed state follows a log-normal distribution with a mean of 5.1 days as described in the Background section.

All cells within the geography are initialized to the same infection state, differing only in population size and their neighborhoods. All age groups of all cells are initialized to 10% Exposed and 90% Susceptible. In the base parameters there are five exposed phases, where $\varepsilon_a(q)$ = [0.2, 0.2, 0.2, 0.2, 0.2]. In the log-normal Exposure simulation there are fourteen Exposed phases, following the profile of Table #3. Although the incubation time of COVID-19 can be up to 24 days [5], 98.5% of people enter infection by day 15, so a T_e of 14 was chosen for simulation efficiency.

X~LogNormal(1.63, 0.5)		
day : x	P(X <= x)	$\boldsymbol{\varepsilon_a(q)}$:P(x > X > x-1)
0	0.0000	
1	0.0007	0.0007
2	0.0349	0.0342
3	0.1580	0.1231
4	0.3345	0.1766
5	0.5075	0.1730
6	0.6493	0.1418
7	0.7555	0.1061
8	0.8312	0.0757
9	0.8838	0.0527
10	0.9200	0.0362
11	0.9447	0.0247
12	0.9616	0.0169
13	0.9732	0.0116
14	0.9812	0.0080

Table 3: Exposed Phase Log N Coefficients

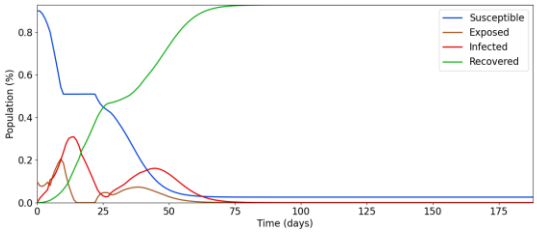

Figure 13: Base SEIRD Ottawa

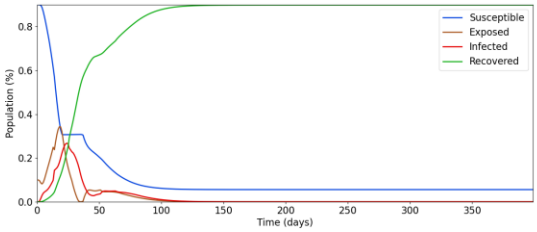

Figure 14: Log-Normal Exposed Ottawa

The base scenario pandemic is more severe, with a higher number of peak Infected and a much stronger second wave of infections post lockdown. It is also a shorter pandemic, reaching steady state after 75 days. The scenario with log-normal incubation time has a higher peak Exposed but a lower peak Infected. This is because the first wave lasts longer because of the later average incubation time. Initializing all cells to the same infection state at the beginning of the simulation is not realistic as infections initially start in with a small amount of infected in a single area. This simulation is included because it is easy for a model user to generate, requiring no manual editing of input JSON data to initialize infected cells.

Next similar scenarios to the previous Ottawa simulations are presented, however the infection now begins in a single centrally located cell with 10% of its entire population as Exposed, and the other 90% as Susceptible. The pandemics start in a centrally located Ottawa cell (DA 35060327) using the log-normal Exposed profile. The model parameters had to be modified to increase virus spread over geography, increasing mobility rates, virulence rates, and disobedience, while decreasing the effectiveness of lockdown (Mobility correction factor). The SEIRD parameters are used are given in Table 4, and the result of a simulation using these parameters are seen below:

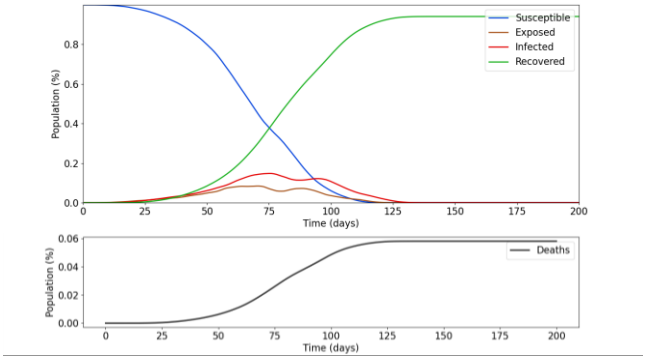

Figure 15: Infection only in DA 35060327

Population Size : **N**	1000
Age Groups : $\mathbf{N_A}$	5
Exposed State Length : $\mathbf{T_e}$	14
Infected State Length : $\mathbf{T_i}$	12
Recovered State Length : $\mathbf{T_r}$	16
Virulence Rates: $\boldsymbol{\lambda_a(q)}$	0.15 for T_i days all ages
Incubation Rates: $\boldsymbol{\varepsilon_a(q)}$	14 day Profile
Recovery Rates: $\boldsymbol{\gamma_a(q)}$	0.07 for T_r days all ages
Mobility Rates: $\mu_a(q)$	1 for T_i days all ages
Fatality Rates: $\mathfrak{f}_a(q)$	0.005 for T_i days all ages
Disobedient: **d**	30%
Hospital Capacity:	0.2
Fatality Modifier: **F**	1.5
Reinfection : **isSIRDS**	false
Correction Factor : $\boldsymbol{m_c}$	"0.2" : [0.2, 0.1]

Table 4: Increased Parameters

Even with the mentioned adjustments, the pandemic never becomes severe. Cells located close to the DA 35060327 may be undergoing lockdowns because of high levels of infection, but the population of the cell space as a whole does not exceed 20% Infected at any time. Much more insight could be gained if the results of this simulation were viewed in a geographical representation like the SIRDS model.

5 Conclusions

The expansion of the Geographical SIRDS Cell-DEVS model to include an Exposed state has been presented. The use of the Cadmium JSON library allowed complex geographical neighborhoods to be generated for use with the model. The addition of the Exposed state behaves intuitively, lengthening the overall pandemic length compared to the SIRD model without an Exposed state. The Exposed state is also seen to worsen peak infections when a lockdown policy would be put into place, as a number of Exposed will still become Infected days after a lockdown policy is put into place.

Many improvements can be made to match this model to the characteristics of COVID-19, including age group specific mobility rates, fatality rates, recovery rates, and virulence rates. The model can be tested against real pandemic data over both small and large geographical areas to verify and tune the model. If this SEIRDS model is used going forward, the GIS Simulation Explorer could be adapted to view simulation data in the geographical regions input to the model [9].

ACKNOWLEDGMENTS

The SIRDS model and as well as the geographical tools used to view results were developed by Binyamin Brion, Román Cárdenas Rodríguez, Gabriel Wainer and Omar Kawach among others.

REFERENCES

[1] "COVID-19 Map," *Johns Hopkins Coronavirus Resource Center*. https://coronavirus.jhu.edu/map.html (accessed Jan. 05, 2021).

[2] *SimulationEverywhere-Models/Geography-Based-Model*. SimulationEverywhere-Models, 2021.

[3] S. Zhong, Q. Huang, and D. Song, "Simulation of the spread of infectious diseases in a geographical environment," *Sci. China Ser. Earth Sci.*, vol. 52, no. 4, pp. 550–561, 2009, doi: 10.1007/s11430-009-0044-9.

[4] L. Anselin, "Review of Spatial Processes, Models and Applications," *Econ. Geogr.*, vol. 59, no. 3, pp. 322–325, 1983, doi: 10.2307/143420.

[5] C. McAloon *et al.*, "Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research," *BMJ Open*, vol. 10, no. 8, p. e039652, Aug. 2020, doi: 10.1136/bmjopen-2020-039652.

[6] N. Waters, "Tobler's First Law of Geography," Dec. 2017, doi: 10.1002/9781118786352.wbieg1011.

[7] B. Zeigler, H. Prähofer, and T. G. Kim, "Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems," vol. 2, Jan. 2000.

[8] G. A. Wainer, "Cellular Modeling with Cell-DEVS: A Discrete-Event Cellular Automata Formalism," in *Cellular Automata*, Cham, 2014, pp. 6–15, doi: 10.1007/978-3-319-11520-7_2.

[9] "staubibr/arslab-web," *GitHub*. https://github.com/staubibr/arslab-web (accessed Jan. 10, 2021).