EPIPOL: An Epidemiological Patterns of Life Simulation (Demonstration Paper)

Will Kohn
will.kohn@emory.edu
Emory University
Atlanta, Georgia, USA

Hossein Amiri
hossein.amiri@emory.edu
Emory University
Atlanta, Georgia, USA

Andreas Züfle
azufle@emory.edu
Emory University
Atlanta, Georgia, USA

ABSTRACT

This paper introduces the EPIPOL disease simulation model, constructed upon the Patterns-of-life simulation, designed to produce human trajectory data. Over recent years, a surge in disease simulation models has been observed, each distinctive in its design and functionality. The primary objective of these models is to predict infection patterns for specified diseases in hypothetical settings, based on all available disease characteristics. A challenge that EPIPOL addresses is the typical rigidity of these models, which are often tailored for a specific disease. Thus they demand profound software expertise for modification. In our demonstration, participants will experience the user-friendliness and clarity of EPIPOL by: (1) selecting disease presets to initialize variables; (2) dynamically adjusting these variables during the simulation for enhanced precision; and (3) visualizing disease propagation in any globally mapped location.

CCS CONCEPTS

• **Information systems** → **Geographic information systems**.

KEYWORDS

Infectious disease simulation, Human behavior, Geospatial simulation, Customization, Epidemic prediction

ACM Reference Format:
Will Kohn, Hossein Amiri, and Andreas Züfle. 2023. EPIPOL: An Epidemiological Patterns of Life Simulation (Demonstration Paper). In *4th ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology (SpatialEpi '23), November 13, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3615898.3628258

1 INTRODUCTION

Gathering real-world data can be expensive and, in many cases, is not feasible. This is why simulations can be the primary source for training machine learning models, especially when predicting hypothetical scenarios or accessing inaccessible real-world data. Infectious diseases are an example of a real-world scenario where data is often inaccessible. While diseases transmit from one person to another, it is often impossible to observe exactly when and between whom an infection event occurred. Even when real-world data is available, it may be highly biased due to factors such as region, culture, language, and more. Specifically for the case of COVID-19 data, a survey of sources of data bias can be found in [6].

Simulating a real-life scenario that accurately captures human life patterns is a daunting task. In such simulations, agents must behave like humans, and crucially, the output data should be both spatial and temporal to emulate high-quality real-world data. Furthermore, the intricacies of life combined with algorithmic complexities make these simulations extremely time-consuming. Simulating life for thousands of agents can be prohibitively expensive, if not outright impossible. The recently published Patterns of Life (POL) simulation [8, 14] is an agent-based simulation that can be run for thousands of agents on various real-world maps and has been used to generate large human mobility datasets [2, 7]. However, it lacks a robust infectious disease model, which is sought-after given the current high demand for medical and infectious disease data. Many contemporary models have evolved to be highly tailored, suitable for a specific disease and setting [11]. While these models fulfill their intended roles by offering realistic projections for the diseases they are crafted for, they often limit users unfamiliar with the software from altering the foundational infrastructure and variables driving the simulation's behavior.

A significant portion of current disease model research focuses intently on epidemiological criteria, seeking to optimize the infection curve to align with real-world observations. In contrast, EPIPOL offers users the flexibility to fine-tune the infection dynamics on-the-fly, providing only foundational guidelines and thresholds to commence. The cornerstone of EPIPOL's prowess is its Patterns-of-life simulation. This simulation takes into consideration the needs of the agents, which consequently influence their movements [8]. Such a design ensures that agents' actions are purposeful, introducing a myriad of adjustable behavioral variables in response to the effects of the disease. This paper presents a detailed demonstration of the EPIPOL disease simulation model, grounded in the Patterns-of-life simulation [8]. Our model leverages a social science-driven simulation rooted in Maslow's hierarchy of needs. This foundation is combined with easily modifiable variables to determine the spread and impact of the disease. An epidemiologist can swiftly adjust the simulation to reflect any global scenario in mere minutes. Collectively, these features underpin the core aim of our research: to develop a disease model with immediate universal adaptability.

The structure of the remainder of this paper is outlined as follows: In Section 2, we delve into the related works. The methodology and modifications to the POL are detailed in Section 3. The demonstrated scenario is detailed in Section 4. Section 5 presents the experimental results, highlighting the impact of varying maps on infectious patterns. Finally, conclusions are drawn in Section 6.

SpatialEpi '23, November 13, 2023, Hamburg, Germany

ACM ISBN 979-8-4007-0360-7/23/11…$15.00
https://doi.org/10.1145/3615898.3628258

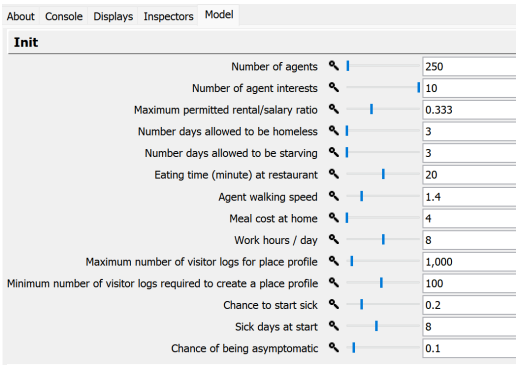

Figure 1: Variable Initialization using the GUI Version

2 RELATED WORKS

Following the widespread outbreak of COVID-19, there has been a surge in research on infectious diseases. Each study leverages various disease models, each offering distinct advantages tailored to specific applications and research objectives. The three primary types of disease models used to model COVID-19 were system dynamics models (SDM), discrete event simulations (DES), and agent-based models (ABM) [13]. SDM are commonly used to understand and predict non-linear behavior over time. They rely on feedback loops and quantitative analysis of various resource flows. Notably, SDM represent the backbone of compartmental disease models [3]. DES are mainly applied to specific locations, such as hospitals, DESs project a chronological sequence of events. Their primary utility stems from optimizing procedures and workflows, thereby enhancing operational efficiency. ABM are stochastic and rely on the movement and interaction of independent agents moving through space. ABMs offer visualization of geospatial phenomena that the two aforementioned models cannot by mapping heterogeneous individuals with distinct characteristics and behaviors, giving a more comprehensive viewing method than a graph or chart. In many of the papers referenced by [13] geospatial data was utilized, that was location-specific and therefore not universally applicable. An exception to this is the Flu and Coronavirus Simulator (FACS) [10]. FACS bears similarities to our developed model but lacks an embedded social sciences framework. Instead of this framework, it evaluates the efficacy of various non-pharmaceutical intervention techniques, akin to the methods described in [1]. FACS incorporates demographic data, which, in a way, acts as an alternative to our social science-based approach. Additionally, FACS accounts for mortality resulting from the disease, an aspect our model currently overlooks.

Numerous papers used the Susceptible, Exposed, Infectious, Recovered (SEIR) model. The research in [12] chose to use an SEIR model, incorporated an ensemble Kalman filter for parameter estimation, and an LSTM to forecast the effects of the disease. The idea of integrating an RNN with a disease simulation model intrigued us. Such an approach could yield more accurate results with reduced computational intensity. However, its accuracy would still hinge on our default ABM, as the LSTM's training would rely on that dataset. Given the prevalence of diseases with a latency period between infection and becoming contagious, we believed that incorporating an SEIR option would enhance our simulation suite's features. This feature will be discussed more in Section 5 and displayed in Figure 2. Our ABM, when leveraged with its capability to adjust both location and disease specifics, presents unique insights. EPIPOL deviates from the intricate mathematical underpinnings that other models heavily depend on. Rather than narrowly concentrating on the ultimate outcomes and implications of the disease, EPIPOL delves deeper into the mechanics of its spread. It further explores how altering various parameters influences the disease's trajectory and ramifications. To our knowledge, EPIPOL stands out as the pioneering ABM that allows customization of the geographical setting while simultaneously employing social sciences to guide agent behavior.

3 METHODOLOGY AND IMPLEMENTATION

We incorporated the infectious diseases model into the Patterns of Life simulation. The source code, along with instructions for running the simulation, can be found at https://github.com/azufle/epipol. To execute the simulation, users have the option of utilizing either the headless version or the Graphical User Interface (GUI) version. For data generation, the headless version is optimal, while the GUI version primarily serves to showcase the simulated world. Figure 1 illustrates the variable initialization within the GUI. Before initiating the simulation, users can select a preset configuration which alters the disease's characteristics. Users have the flexibility to modify these characteristics during the simulation's progression. For our results collection, default properties were employed. When starting the simulation, but before activating the play button, users can adjust any variable of their choice under the 'Model' tab, with important adjustments found in the 'Init' section. We note that these variables can only be modified before the simulation commences and are invoked solely during the model's initialization.

4 DEMONSTRATED SCENARIO

For the demonstration to be presented at SpatialEpi'23, we will use default settings of a generic airborne infectious disease. To highlight the 'change location' feature, we allow users to select between different maps, including Atlanta, GA, USA; Berlin, Germany; Crawfordville, FL, USA; and Aldie, VA, USA. The user may then initiate the simulation by clicking the 'play' button on the GUI version. The GUI will show the movement of agents as well as the resulting epidemiological curves as shown in Figure 2. After running an initial default simulation, users may change the variables of the simulation under the model tab of the GUI. Users can select pre-defined ***.properties*** files to load using predefined infectious disease scenarios. The simulation includes three prebuilt property sets: the default properties (which describe a generic airborne infectious disease), ***influenza.properties*** and ***rhinovirus.properties***. It's crucial to note that these property sets serve as simple models of complex infectious disease spread and their spread patterns may not accurately reflect future infectious disease in real-world.

Users may also change parameters that define whether and how susceptible, exposed, infected, and recovered agents change their respective behavior. For example, we can impose a stay-at-home order for all infectious agents having 80% of the agents obey this order. We can also implement a change in behavior for agents that are not infected yet for self-preservation by leveraging existing work on infectious disease-related change in human mobility [5].

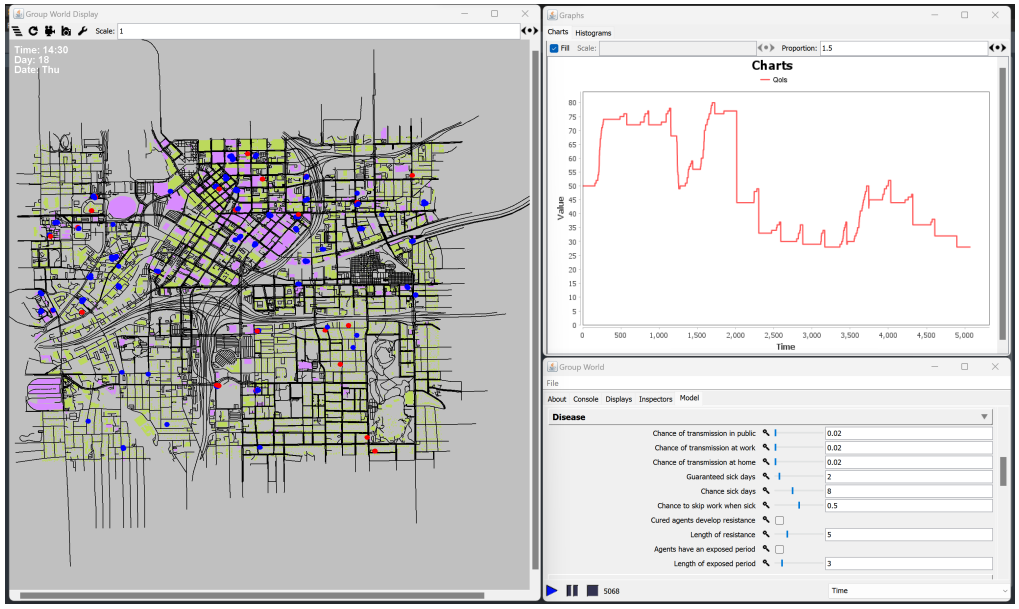

Figure 2: Simulation GUI for the city of Atlanta, Georgia USA

Figure 2 shows a screenshot of the demonstrator using the map of Atlanta, GA, USA. This map excludes its suburbs. Agents in blue represent healthy individuals. Those in red are infected and contagious. By default, these agents interact through their daily routines, transmitting the disease to one another. An available toggle can introduce disease immunity to agents who were previously infected and have since recovered; these agents are shown in green. Agents who have been exposed but aren't contagious will appear purple when this feature is activated. Most of the epidemiological customization can be found in the disease section of the console. The potency and behavior of the disease can be modified during the simulation. Achieving a realistic curve by fine-tuning the disease can be challenging. However, the user can utilize both the map and graphs to visualize the behavior of their disease throughout the simulation. Additionally, there's a default 10% chance of an agent being asymptomatic; however, this can be adjusted in the 'Init' section. Asymptomatic agents do not exhibit any behavioral changes resulting from infection. The flow chart detailing potential agent statuses, when the SEIR model is enabled, is found in Figure 3.

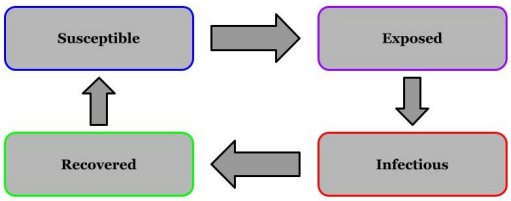

Figure 3: Flowchart illustrating the various states of an agent during the infection procedure.

5 EXPERIMENTAL RESULTS

To showcase the capabilities of the simulation, we compared the spread of a disease across multiple cities. Due to the computational demands of simulating an entire major city, we limited our focus to their downtown sections. Moreover, we sought to evaluate our simulation's accuracy of disease spread in rural areas, a known limitation of many current models [4].

We performed the experiments in two iterations: first, using a constant number of agents, and second, with a number normalized to reflect the actual population density of the cities. In the first iteration, we set the number of agents at 250. Each city was subjected to an infection based on the default disease properties, as defined in the ***parameters.properties*** file.

With a constant number of agents, the results showed minimal variation among the four simulated cities. We selected Atlanta and Berlin as representations of urban areas, while Aldie, VA, and Crawfordville, GA served as our rural choices. Each city underwent ten simulations, and the outcomes were then averaged.

Regardless of the presentation format, the data does not reveal a marked difference between rural and urban areas. This neither confirms nor refutes the accuracy of our simulations, especially since there isn't detailed real-world data available for a direct comparison. The primary source of inaccuracy is the uniform simulated population of 250. While this number is negligible for cities like Atlanta and Berlin, it surpasses half of Crawfordville's population (481) and exceeds the entire population of Aldie (235). When we executed the simulation using populations proportional to the actual sizes of the cities, the results were more definitive. Still, each city was simulated ten times, and the outcomes were averaged.

The comparison of disease spread patterns across different cities, based on relative population sizes, is illustrated in Figure 4. To determine the number of agents for our simulations, we referred to the actual populations of Crawfordville (481) and Aldie (235). For

$\alpha = 0.05$	df=100	n=100
Mean difference	-0.045	
95 percent confidence interval	(-0.050,	-0.040)
p-value	< 2.2e-16	

Table 1: Rural vs urban paired t test

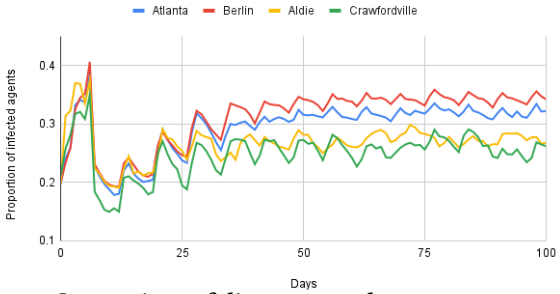

Figure 4: Comparison of disease spread patterns across various cities using relative population sizes.

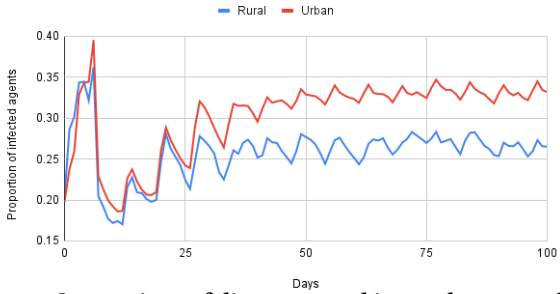

Figure 5: Comparison of disease spread in rural versus urban areas using relative population sizes.

the urban areas, we calculated the number of agents by multiplying the cities' respective population densities with the square mileage of their maps, as obtained from QGIS. This approach yielded 22,479 agents for Atlanta and 61,309 for Berlin. However, simulating these numbers, even in headless mode, was computationally expensive. Consequently, we scaled down by a factor of ten, leading to 2,248 agents for Atlanta and 6,131 for Berlin. Using the actual agent counts would likely have resulted in even more pronounced results. The sharp increase up to day eight, followed by a subsequent decline, is attributed to the consistent number of sick days for agents who began the simulation already infected.

The disease spread comparison between rural and urban areas, based on relative population sizes, is illustrated in Figure 5. Observing a distinct difference between the rural and urban data graphically, and with all criteria satisfied, we proceeded with a paired t-test. The subsequent results are presented in Table 1.

Given that $p < 0.05$, we can reject the null hypothesis. This indicates a statistically significant difference in the proportion of infected agents between rural and urban areas in our simulation. Notably, when the number of agents is adjusted to reflect population density, the proportion of infected agents in urban areas exceeds that in rural areas.

6 CONCLUSION

We introduce a user-friendly, social science-driven disease simulation model tailored for customization to specific user requirements. This simulation offers a visual representation of a disease's hypothetical spread worldwide, provided there is sufficient data on both the disease and the geographical mapping of the area. In the demonstration, users can craft a unique disease in a chosen location using the console depicted in Figure 1 and Figure 2 . We envision that EPIPOL will be leveraged by professional epidemiologists to anticipate potential epidemic impacts, thereby enhancing our response measures. In future iterations, we aim to broaden the range of agent behavioral modifications during infection, building on our current model where the sole change is the option for agents to skip work when ill [9]. The foundation for this is already laid out in the Patterns-of-life simulation, making integration straightforward. We also plan to introduce toggles for different infection duration distributions to enhance the realism of our model. At present, the model only supports a uniform distribution. Introducing additional distributions such as normal, Poisson, and binomial will be a priority. Moreover, simplifying the map generation process is on our radar to expedite the deployment process. Once these enhancements are in place, we look forward to sharing the updated version with both the public and the scientific community.

7 ACKNOWLEDGEMENTS

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/ Interior Business Center (DOI/IBC) contract number 140D0423C0025. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI/IBC, or the U.S. Government.

REFERENCES

[1] Shubhada Agrawal, Siddharth Bhandari, Anirban Bhattacharjee, et al. 2020. City-scale agent-based simulators for the study of non-pharmaceutical interventions in the context of the COVID-19 epidemic: Iisc-tifr covid-19 city-scale simulation team. *Journal of the Indian Institute of Science* (2020).

[2] Hossein Amiri, Shiyang Ruan, Joon-Seok Kim, Hyunjee Jin, Hamdi Kavak, Andrew Crooks, Dieter Pfoser, Carola Wenk, and Andreas Züfle. 2023. Massive Trajectory Data Based on Patterns of Life (Data and Resources Paper). In *31st ACM SIGSPATIAL Int. Conference on Advances in Geographic Information Systems.*

[3] Raul Bagni, Roberto Berchi, and Pasquale Cariello. 2002. A comparison of simulation models applied to epidemics. *Journal of Artificial Societies and Social Simulation* (Jun 2002).

[4] Todd Easton, Kyle Carlyle, Joseph Anderson, and Matthew James. 2011. Simulating the Spread of an Epidemic in a Small Rural Kansas Town. *Int. J. Artif. Life Res.* 2, 2 (apr 2011), 95–104.

[5] Song Gao, Jinmeng Rao, Yuhao Kang, Yunlei Liang, and Jake Kruse. 2020. Mapping county-level mobility pattern changes in the United States in response to COVID-19. *SIGSpatial Special* 12, 1 (2020), 16–26.

[6] Gareth J Griffith, Tim T Morris, Matthew J Tudball, et al. 2020. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nature communications* 11, 1 (2020), 5749.

[7] Joon-Seok Kim, Hyunjee Jin, Hamdi Kavak, Ovi Chris Rouly, Andrew Crooks, Dieter Pfoser, Carola Wenk, and Andreas Züfle. 2020. Location-based social network data generation based on patterns of life. In *2020 21st IEEE International Conference on Mobile Data Management (MDM).* IEEE, 158–167.

[8] Joon-Seok Kim, Hamdi Kavak, Umar Manzoor, Andrew Crooks, Dieter Pfoser, Carola Wenk, and Andreas Züfle. 2019. Simulating urban patterns of life: A geo-social data generation framework. In *Proceedings of the 27th ACM SIGSPATIAL international conference on advances in geographic information systems.* 576–579.

[9] Jongmin Lee, Seok-Min Lee, and Eunok Jung. 2021. How Important Is Behavioral Change during the Early Stages of the COVID-19 Pandemic? A Mathematical Modeling Study. *International Journal of Environmental Research and Public Health* 18, 18 (Sep 2021), 9855.

[10] Imran Mahmood, Hamid Arabnejad, Diana Suleimenova, et al. 2022. FACS: A geospatial agent-based simulator for analysing COVID-19 spread and public health measures on local regions. *Journal of Simulation* 16, 4 (2022), 355–373.

[11] Feres A. Salem and Ubirajara F. Moreno. 2022. A multi-agent-based simulation model for the spreading of diseases through social interactions during pandemics. *Journal of Control, Automation and Electrical Systems* 33, 4 (2022), 1161–1176.

[12] Yifan Yang, Wenwu Yu, and Duxin Chen. 2020. Prediction of COVID-19 spread via LSTM and the deterministic SEIR model. In *39th Chinese Control Conf.* 782–785.

[13] Weiwei Zhang, Shiyong Liu, Nathaniel Osgood, et al. 2022. Using simulation modelling and systems science to help contain COVID-19: A systematic review. *Systems research and behavioral science* (Aug 2022).

[14] Andreas Züfle, Carola Wenk, Dieter Pfoser, Andrew Crooks, Joon-Seok Kim, Hamdi Kavak, Umar Manzoor, and Hyunjee Jin. 2023. Urban life: a model of people and places. *Comp. and Math. Organization Theory* 29, 1 (2023), 20–51.