

Spatial modeling of influenza outbreaks in Saint Petersburg using synthetic populations [★]

Vasily Leonenko¹, Alexander Lobachev¹, and Georgiy Bobashev²

¹ ITMO University, 49 Kronverksky Pr. Saint Petersburg, 197101, Russia

² RTI International, 3040 Cornwallis Rd. P.O. Box 12194 NC 27709

Abstract. In this paper, we model influenza propagation in the Russian setting using a spatially explicit model and a detailed human agent database as its input. The aim of the research is to assess the applicability of this modeling method using influenza incidence data for 2010-2011 epidemic outbreak in Saint Petersburg and to compare the simulation results with the output of the compartmental SEIR model for the same outbreak. For this purpose, a synthetic population of Saint Petersburg was built and used for the simulation via FRED open source modeling framework. The parameters related to the outbreak (background immunity level and effective contact rate) are assessed by calibrating the compartmental model to incidence data. We show that the current version of synthetic population allows the agent-based model to reproduce real disease incidence.

Keywords: epidemiology · synthetic populations · seasonal influenza · agent-based modeling · FRED.

1 Introduction

Today 55% of world's population lives in cities, and this number, according to UN predictions, is expected to reach 68% by 2050 [31]. Thus, the importance of cities for human societies is increasing over time. Due to their intricate structure, modern cities constitute a perfect example of complex systems, and our ability to understand them scientifically is limited [5]. So is the situation with the social and economic processes within them.

One of the processes intrinsically connected with urban structure is influenza epidemics. Seasonal influenza causes repetitive epidemic outbreaks resulting in high worker/school absenteeism, productivity losses and death cases due to disease complications. According to WHO [33], the corresponding annual number of mortality cases reaches whopping 500 thousand. To anticipate the incoming outbreaks and prepare healthcare infrastructure to fight with their detrimental ramifications, statistical and mathematical models are widely used. The predictive force of these models is limited by the fact that the influenza outbreak

[★] This work is financially supported by Ministry of Education and Science of the Russian Federation, Agreement #14.575.21.0161 (26/09/2017). Unique Identification RFMEFI57517X0161.

dynamics in urban settings is driven by a lot of factors, some of which are hard to be quantified. Along with the weather factors [29], [20], [27], the important role is played by the behavior of the human population, such as daily migration patterns resulting in different effective contact probabilities [6], [30]. In fact, the bigger a city is, the milder is the response of the disease incidence to climate forcing and the more important the human-related factors become [8]. These factors also deserve attention due to occurrence of background immunity against influenza, which is a result of repetitive flu outbreaks caused by similar flu strains [17]. It might be assumed that peculiarities of commuting patterns of citizens and geographical distribution of their dwellings subsequently cause different distribution of effective contacts between the susceptibles and the infectives, leading to changes in immunity levels of the individuals in different cities. For instance, a highly connected city, where mass action law assumption [34] generally holds, might have an epidemic dynamics and consequently a distribution of the immune quite different from the city with apparent geographical clustering. The accumulation of these differences due to faster circulation of flu virus around the globe might be the reason of the failure of the approach which was earlier used to predict flu epidemics in Soviet Union [15]. The mentioned approach was based on the assumption that a forthcoming influenza outbreak dynamics could be predicted using the data from the cities already affected by the epidemic during the season under consideration, which worked in 1970s, but is not true anymore [22]. Modeling flu propagation using detailed population structures which (somewhat) accurately reflect the peculiarities of urban contact patterns might allow us to quantify the role of contact patterns on the formation of background immunity and to assess how the differences in city structures lead to different flu epidemic dynamics. This paper is considered to be the first step in the stated direction.

The aim of this work is to create a detailed description of urban population in the Russian setting and couple it with agent-based modeling framework to perform a simulation of flu dynamics. Using our previous results obtained in the field of flu outbreak modeling in Russia [19], [21], we want to compare the output of a spatially explicit model with the one of standard SEIR compartmental model of Kermack–McKendrick type and to demonstrate the ability of the former to produce more plausible results than the latter. For this purpose, we regard influenza outbreak in Saint Petersburg in 2010–2011 as a case study. The city was chosen due to large populace (it’s the second largest city in Russia), economic and cultural importance, and abundance of detailed data on influenza incidence (the records are available from 1935).

2 Synthetic population

“Synthetic population” is a synthesized, spatially explicit human agent database (essentially, a simulated census) representing the population of a city, region or country. By its cumulative characteristics, this database is equivalent to the real population but its records does not correspond to real people – this fact helps

avoid privacy issues. In our research, we employed the approach for synthetic population generation developed by RTI International [32], which was used by various research groups to create populations for 50 US states, along with another regions and countries. Statistical and mechanistic models built on top of the synthetic populations helped tackle a variety of research problems, including those connected with public health. Statistical analysis of opioid-related overdoses in Cincinnati [4] can be named as an example.

According to the standard of RTI International, a synthetic population consists of several txt-files, each of them containing a table with every row being a single record corresponding to some entity – an individual, a household, a workplace, a school, etc. The full list of files with their short description is presented in Table 1. Sticking to the same standard, we generated the files corresponding to the population of Saint Petersburg. Since the data available for Saint Petersburg was not complete, we altered or omitted some of the methods, resulting in a simplified population, which, however, seems to satisfy our demands related to influenza modeling. The details of input data we used and the algorithms we employed to generate the population follow.

2.1 Household data

The principal data source for our synthetic population is 2010 data from “Edinaya sistema ucheta naseleniya Sankt Peterburga” (“Unified population accounting system of Saint Petersburg”) [11]. The data is represented in a form of Excel spreadsheets containing records with house addresses and the corresponding number of dwellers of certain age and gender (see Table 2).

To match the household addresses with the geographical coordinates and assess the plausibility of the obtained geographical data, a computational algorithm was developed and implemented using Python programming language. The details of the algorithm implementation follow.

Adding object coordinates.

- For each record:
 - Form the address string using the information from the address fields of the record in the format “city” + “street” + “house”.
 - Feed the address string to Yandex.Geocoder online service [36] which returns the latitude and the longitude of the object by this address.
 - Add coordinates to a record.

Removing implausible data. Since the record addresses were apparently derived from handwritten data or manually typed, in some cases they are incomplete or contain typos. The geocoder we used always returns two coordinates as an output, no matter whether he processed the input successfully or not. If an address is not interpreted correctly, Yandex.Geocoder makes guesses on what the correct address should be, which often results in semi-random coordinates. We use an empirical algorithm to filter out those obviously senseless results.

File	Contents
Files used in the current population version	
<code>households.txt</code>	Contains the location and descriptive attributes for each household. Household records in the <code>households.txt</code> file link to individual person records in the <code>people.txt</code> table.
<code>people.txt</code>	Contains a record for each person, along with his or her age and sex. These synthetic person records link to the <code>households.txt</code> file (via the <code>sp_hh_id</code> field)
<code>schools.txt</code>	Contains a record for each school, along with its zip code, maximum capacity and coordinates
<code>workplaces.txt</code>	Contains a record for each workplace, along with its coordinates and size
Empty or omitted files	
<code>hospitals.txt</code>	Contains a record for each hospital, along with its coordinates, number of physicians and beds. Contains zero records in this version of the synthetic population.
<code>gq.txt</code>	Contains a record for each general quarters, along with their type (prisons, student dorms, etc), coordinates and capacity. Contains zero records in this version of the synthetic population.
<code>gq_people.txt</code>	Contains a record for each person, which lives in general quarters, along with his or her age and sex. These synthetic person records link to the <code>gq.txt</code> file (via the <code>sp_gq_id</code> field). Contains zero records in this version.
<code>pums.p.txt</code>	Contains personal records from the public use microdata series. Links to the <code>people.txt</code> file the <code>serialno</code> field. Absent in this version of the synthetic population.
<code>pums.h.txt</code>	Contains household records from the public use microdata series. Links to the <code>households.txt</code> file the <code>serialno</code> field. Absent in this version of the synthetic population.

Table 1. File structure of a synthetic population for Saint Petersburg.

For this procedure, we rely on a number of empirical assumptions related to matches between the location (coordinate) and the text address. E.g., if there are multiple addresses to which only one pair of coordinates is assigned, we remove all such records from the database except the first one, summing the corresponding numbers of dwellers.

2.2 School data

The list of schools and their addresses was formed manually using the data from the official web-site of the Government of Saint Petersburg [12]. The coordinates of schools were found using Yandex.Geocoder in the same fashion, as it was done for dwellings.

Column names	Contents
city, street, house	Dwelling address
g0, g1 ... g100	Number of female dwellers of ages 0, 1, ..., 100
m0, m1 ... m100	Number of male dwellers of ages 0, 1, ..., 100

Table 2. Data source format.

2.3 Workplace data

The distribution of working places for adults and their coordinates were derived from the data obtained with the help of Yandex.Auditorii API [35]. Initially the data was available in a form of a .geojson file which consisted of relative workplace size assessments for each of the cells in a hexagonal grid (see Figure 1). This data was normalized using the official cumulative employment numbers [10]. Synthetic workplace records were created by assigning the calculated number of employees in each hexagonal cell to imaginary geographical location coinciding with the center of this cell.

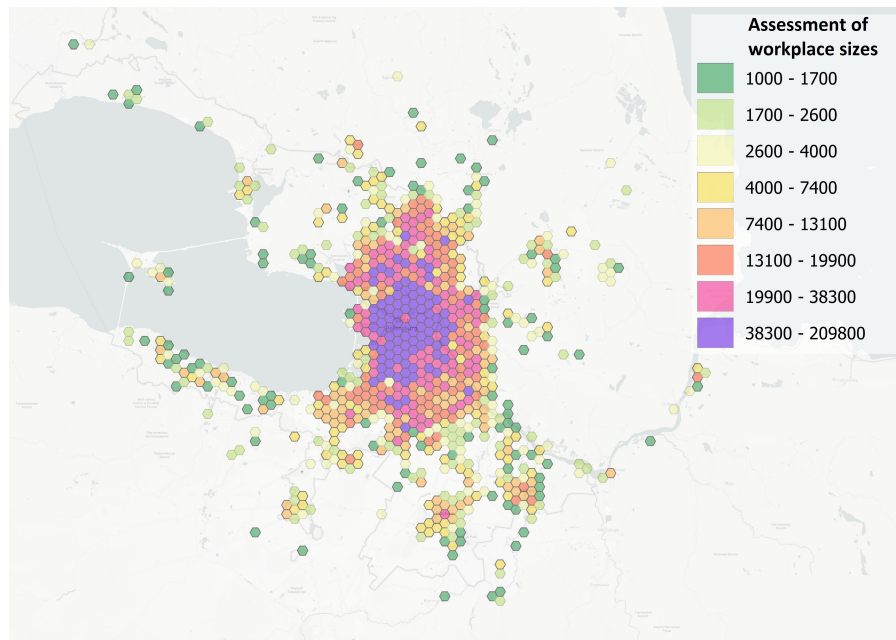


Fig. 1. The distribution of working places in St Petersburg. The numbers are given before the normalization.

2.4 Assigning people to schools and workplaces

We assumed that young people aged 7 to 17 attend schools, and the adults of working age (18 to 55 for males and 18 to 60 for females) might be working. Iterating through the list of records in `people.txt`, we were assigning each person to a closest school or working place, until they are filled to capacity or there is no more people to be assigned.

3 Agent-based modeling

An open-source framework FRED [13], [26] was used for the simulations. The framework has discrete time, with the modeling step equal to one day. The epidemic process is initiated by assigning randomly an infectious status to some individuals in the population at the beginning of the simulation. In addition to that, the infection can be seeded according to a user-specified schedule, reflecting the external infection process.

The contacts among the individuals that lead to new infection cases are modeled in the following way.

- Each agent in the population potentially interacts with other agents with whom he shares activity locations. These locations include schools, workplaces, households and home neighborhoods (defined as 1 km square cells around the agent’s household).
- During the weekends, schools are considered to be closed and most workers equally do not attend their working places. At the same time, the number of neighborhood contacts increases by 50%.
- The rate of effective contacts in a particular activity location depends on the expected number of contacts per infectious person per day and the infection transmission probability. The expected number of contacts is considered not to be dependent on the place size.

The output of the framework in a form of csv-files contains quantities and spatial distributions of individuals of four groups (susceptible, exposed, infected, recovered) at every time step. We used Python programming language and QGIS open source software to process the results and create maps and incidence graphs. An example of the map of influenza propagation is shown in Figure 2.

3.1 Influenza incidence data

The original dataset provided by the Research Institute of Influenza [1] contains cumulative weekly incidence, i.e., the number of new acute respiratory infection (ARI) cases per day in Russian cities, which includes influenza and other respiratory infections. Before the model fitting, we had to refine the incidence data by restoring the missed values and correcting the under-reporting biases. We also needed to extract flu incidence from the cumulative ARI incidence data. Corresponding algorithms are described in detail in [20], here we introduce briefly the sequence of operations.

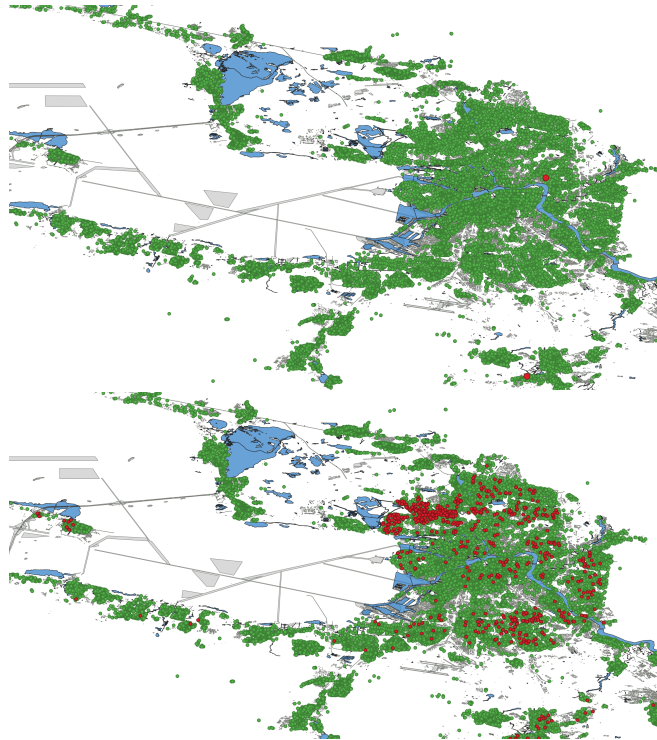


Fig. 2. The distribution of disease-free (green) and infected (red) households in Saint Petersburg: (a) Day 1, (b) Day 8.

- **Under-reporting correction.** Since infected people avoid visiting health-care facilities during holidays, the corresponding weekly prevalence is lower than the actual number of newly infected. This under-reporting bias can be corrected by means of cubic interpolation [3] using the incidence registered in the adjacent weeks.
- **Bringing the incidence data to daily format.** Daily incidence is obtained with the help of cubic interpolation of weekly incidence. We assume that $n_{inf}^{Thu} = n_{inf}^W / 7$, where n_{inf}^W is the weekly incidence taken from the database and n_{inf}^{Thu} is the daily incidence for Thursday of the corresponding week.
- **Extracting incidence data related to influenza outbreak.** At first, the algorithm finds higher non-influenza ARI incidence level, which corresponds to average daily number of newly infected during the months when influenza might occur in temperate regions (figure 3, red horizontal dashed line). The part of the graph, which is attributed to a flu outbreak (figure 3, red solid line), should have its peak well above the higher ARI level. It should also comply with the time period during which the ARI prevalence exceeds the

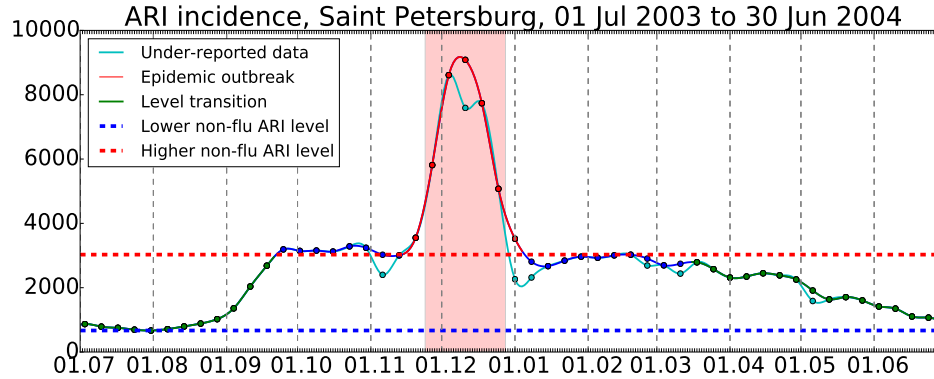


Fig. 3. Influenza outbreak data extraction from the interpolated ARI incidence.

non-epidemic ARI threshold assessed in the Flu Research Institute (figure 3, red rectangle). The beginning and ending of the extracted curve is chosen to match the higher ARI incidence level. The first incidence point of the curve is considered to be the first day of the epidemic outbreak.

3.2 Fitting a SEIR model to data

The model we use for fitting is a standard SEIR compartmental model in a form of a system of ordinary differential equations (see [19] for the details). Let $Z^{(dat)}$ be the set of incidence data points loaded from the input file and corresponding to one particular outbreak. Assume that the number of points is t_1 , which equals the observed duration of the outbreak. The fitting algorithm selects the values of model parameters corresponding to the model output which minimizes the distance between the modeled and real incidence points:

$$F(Z^{(mod)}, Z^{(dat)}) = \sum_{i=0}^{t_1} (z_i^{(mod)} - z_i^{(dat)})^2,$$

Here $z_i^{(dat)}$ and $z_i^{(mod)}$ are the absolute incidence numbers for the i -th day taken from the input dataset and derived from the model correspondingly. The limited-memory BFGS optimization method is used to find the best fit [24]. Since the existence of several local minima is possible, the algorithm has to be launched several times with different initial values of input variables. The best fit is chosen as a minimum among the distances achieved from all the algorithm runs. To characterize the goodness of fit we utilize the coefficient of determination $R^2 \leq 1$. This coefficient shows the fraction of the response variable variation that is explained by a model [25]. The detailed description of the fitting procedure is available in [19], [21].

3.3 Simulation

By fitting the SEIR model to ARI incidence for influenza outbreak in St Petersburg in 2010-2011 (see Figure 5a), we assess two parameters: the background immunity level $1 - \alpha$ and the effective contact intensity λ . The detailed data on the distribution of dwellers we used gave slightly less people in total than it was claimed by official statistics, so we normalized the susceptible ratio α to account for this. The obtained parameter values were used in FRED simulation along with the default values for influenza epidemics provided with the framework (see Table 3). The overall scheme of the described process is presented in Fig. 4.

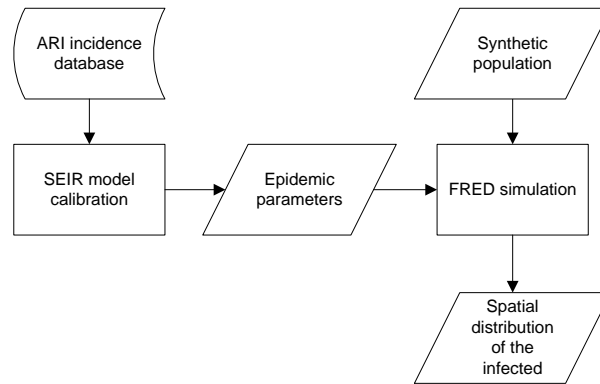


Fig. 4. FRED simulation scheme

The simulation was run 100 times with different seed values for the random number generator. Influenza incidence data obtained as a result of each simulation run was used to find confidence intervals for the daily influenza incidence — see Figure 5b.

As Figure 5 shows, despite the imminent uncertainty in synthetic population data, the output of the agent-based model demonstrates a satisfactory agreement with actual ARI incidence. Also, apart from the compartmental SEIR model, spatially explicit simulation demonstrates the right-skewed incidence curve, which better conforms to ARI data. This form of the curve might reflect different number of social connections between the individuals. The so called “superspreaders” are infected first and cause numerous infection cases, whereas the less socialized persons got reached by the flu later and contribute somewhat to infection process, making the decline of disease incidence slower. The value $1 - \alpha$ of background immunity level used in both models leads to almost the same peak heights (see Table 4).

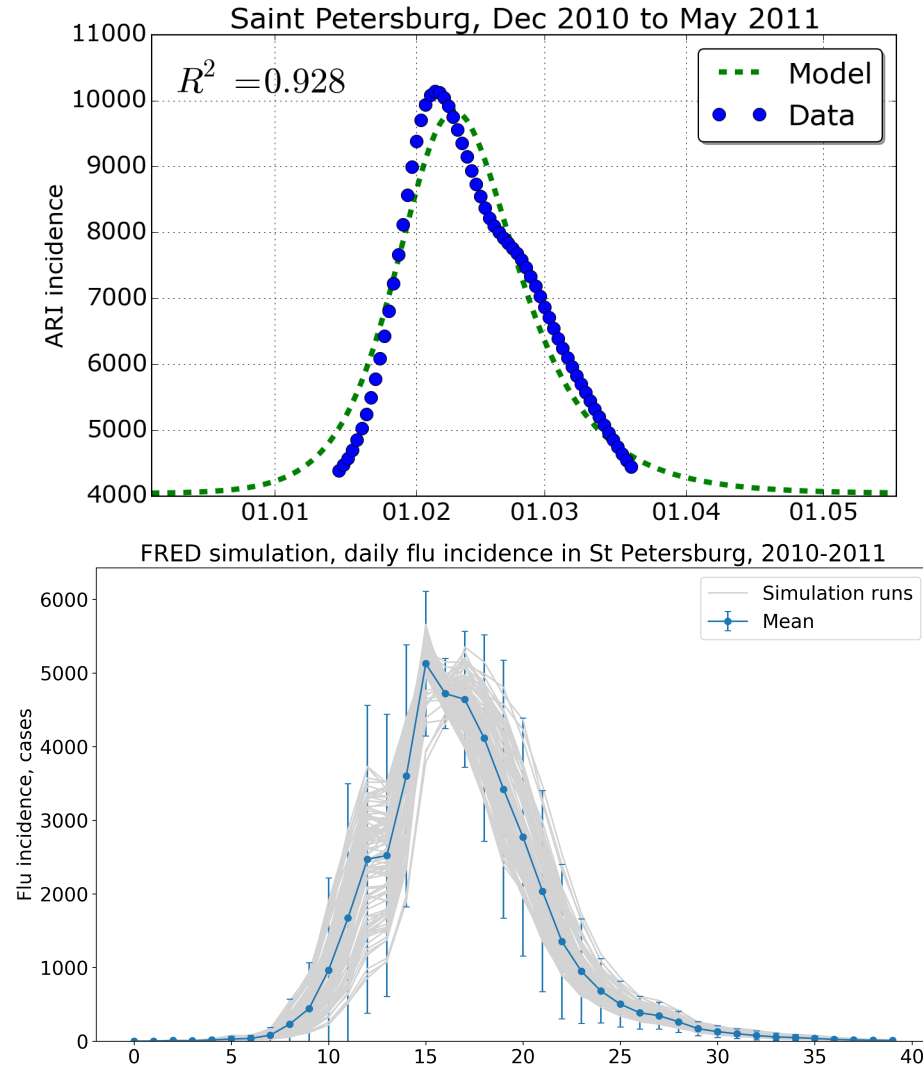


Fig. 5. (a) SEIR model calibration; (b) FRED output generated using the obtained parameter values. Note that the first graph demonstrates the cumulative ARI incidence (influenza and other acute respiratory diseases), whereas the second one shows disease incidence attributed solely to influenza, thus the baselines on the graphs correspond to a seasonal ARI level in the first case (around 4000 newly infected per week), and zero in the second case.

Parameter name	Description	Value	Source
INF.transmissibility	A coefficient that modulates the transmissibility of a condition	6.55	Estimated
INF.S.susceptibility	A ratio of susceptible individuals in the population	0.46	Estimated
INF.E.duration_distribution	Distribution of exposed state duration	lognormal	Default
INF.E.duration_median	Median of exposed state duration	1.9	Default
INF.E.duration_dispersion	Dispersion of exposed state duration	1.23	Default
INF.Is.duration_distribution	Distribution of infected state duration	lognormal	Default
INF.Is.duration_median	Median of infected state duration	5.0	Default
INF.Is.duration_dispersion	Dispersion of infected state duration	1.5	Default

Table 3. Model parameters ('INF' stands for 'influenza', 'Is' is 'infectious symptomatic')

Parameter name	Data	Simulation
Outbreak duration, days	65	51
Maximum incidence day (from the outbreak onset)	22	15
Maximum incidence height, cases per day	5756	5126

Table 4. The incidence curve parameters obtained in the simulation compared to those of the real outbreak

4 Discussion and future work

As we showed in this research, coupling of synthetic populations with agent-based models is a feasible approach which allows to perform spatially explicit influenza propagation modeling in Russian settings, even when a limited number of data is used to reconstruct the urban population. The apparent drawback of the current synthetic population is the procedure of assigning people to schools and workplaces, described in Section 2.4. The idea of picking the closest available spot in a school/workplace was favored for its simplicity, but it is obviously unrealistic. We plan to create several populations corresponding to different assigning algorithms (from picking a workplace/school at random within the whole city to seeking it in a limited radius from the person's home location) and to compare the results of the simulation runs. By doing that, we expect to quantify the variance in outbreak duration, peak day and peak height related to contact networks of different topological structures.

Another drawback of the performed simulation lies in the approximate nature of deriving disease-related parameters for FRED. Particularly, we cannot exactly match the value of `INF.transmissibility` to value of λ from the compartmental model, because they are not equivalent, although correlated. To obtain more realistic disease incidence generated by the model, global optimization techniques will be used by the authors in the same fashion as it was done earlier for the compartmental SEIR models [28]. Since repetitive simulations with different input parameter sets are computationally expensive, we consider applying methods for assessing and reducing the uncertainty in disease-related input parameters [2] which will decrease the state space of the model, and implementing an agent-based model using general-purpose computing on graphics processing units [23] to achieve a speedup.

In the current research we did not consider contacts in public transport. Although the research on the role of New York subway in disseminating influenza showed that its effect is slight [7], we still find it necessary to question this conclusion, because Russian commute patterns and the nomenclature of social groups which use metro may differ from the one of New York.

We assume that, in the long run, the influenza propagation modeling using synthetic populations will allow us to:

- classify the cities into several groups, depending on their geographical structure, contact pattern types and, subsequently, peculiarities of influenza dynamics and use models calibrated to one cities to predict epidemics in the others within the same group;
- reconstruct the pattern of background immunity formation as a function of urban and epidemic factors.

As a byproduct, synthetic populations created for the cities under consideration will be freely available and might facilitate solving urban issues with the help of modeling. The immediate gain we expect from the created synthetic population for Saint Petersburg is that research groups within our department, which work with problems such as transport planning [16], [18], ambulance dispatching [9] and crime rate assessment [14], might want to switch from gathering data from scratch for every particular statistical or agent-based model to using a unified population. Since the synthetic population conforms to a certain defined standard, it might be easily reused in different projects and elaborated further by mutual efforts to everyone's benefit.

References

1. Research Institute of Influenza website. [online], <http://influenza.spb.ru/en/>
2. Artzrouni, M., Leonenko, V.N., Mara, T.A.: A syringe-sharing model for the spread of HIV: application to Omsk, Western Siberia. *Mathematical medicine and biology: a journal of the IMA* **34**(1), 15–37 (2015)
3. Baroyan, O., Basilevsky, U., Ermakov, V., Frank, K., Rvachev, L., Shashkov, V.: Computer modelling of influenza epidemics for large-scale systems of cities and

- territories. In: Proc. WHO Symposium on Quantitative Epidemiology, Moscow (1970)
4. Bates, S., Leonenko, V., Rineer, J., Bobashev, G.: Using synthetic populations to understand geospatial patterns in opioid related overdose and predicted opioid misuse. *Computational and Mathematical Organization Theory* pp. 1–12 (2018)
 5. Bettencourt, L.M.: The origins of scaling in cities. *science* **340**(6139), 1438–1441 (2013)
 6. Cauchemez, S., Valleron, A.J., Boelle, P.Y., Flahault, A., Ferguson, N.M.: Estimating the impact of school closure on influenza transmission from sentinel data. *Nature* **452**(7188), 750 (2008)
 7. Cooley, P., Brown, S., Cajka, J., Chasteen, B., Ganapathi, L., Grefenstette, J., Hollingsworth, C.R., Lee, B.Y., Levine, B., Wheaton, W.D., et al.: The role of subway travel in an influenza epidemic: a new york city simulation. *Journal of Urban Health* **88**(5), 982 (2011)
 8. Dalziel, B.D., Kissler, S., Gog, J.R., Viboud, C., Bjørnstad, O.N., Metcalf, C.J.E., Grenfell, B.T.: Urbanization and humidity shape the intensity of influenza epidemics in us cities. *Science* **362**(6410), 75–79 (2018)
 9. Fu, X., Presbitero, A., Kovalchuk, S.V., Krzhizhanovskaya, V.V.: Coupling game theory and discrete-event simulation for model-based ambulance dispatching. *Procedia Computer Science* **136**, 398–407 (2018)
 10. Government of Saint Petersburg: Edinaya sistema ucheta naseleniya Sankt Peterburga (Unified population accounting system of Saint Petersburg). [online], <https://reestr-gis.spb.ru> (In Russian)
 11. Government of Saint Petersburg: Labor and employment committee. information on economical and social progress. [online], <http://rspb.ru/analiticheskaya-informaciya/razvitie-ekonomiki-i-socialnoj-sfery-sankt-peterburga/> (In Russian)
 12. Government of Saint Petersburg: Official web-site. [online], <https://www.gov.spb.ru/>
 13. Grefenstette, J.J., Brown, S.T., Rosenfeld, R., DePasse, J., Stone, N.T., Cooley, P.C., Wheaton, W.D., Fyshe, A., Galloway, D.D., Sriram, A., et al.: FRED (a framework for reconstructing epidemic dynamics): an open-source software system for modeling infectious diseases and control strategies using census-based populations. *BMC public health* **13**(1), 940 (2013)
 14. Ingilevich, V., Ivanov, S.: Crime rate prediction in the urban environment using social factors. *Procedia Computer Science* **136**, 472–478 (2018)
 15. Ivannikov, Y., Ogarkov, P.: An experience of mathematical computing forecasting of the influenza epidemics for big territory. *Journal of Infectology* **4**(3), 101–106 (2012), In Russian.
 16. Khodnenko, I., Kudinov, S., Smirnov, E.: Walking distance estimation using multi-agent simulation of pedestrian flows. *Procedia Computer Science* **136**, 489–498 (2018)
 17. Konshina, O., Sominina, A., Smorodintseva, E., Stolyarov, K., Nikonorov, I.: Population immunity to influenza virus a(h1n1)pdm09, a(h3n2) and b in the adult population of the Russian federation long-term research results. *Russian Journal of Infection and Immunity* **7**(1), 27–33 (2017). <https://doi.org/10.15789/2220-7619-2017-1-27-33>, in Russian.
 18. Lantseva, A.A., Ivanov, S.V.: Assessment of pedestrian flow volumes through public transport modelling. *Procedia Computer Science* **136**, 463–471 (2018)
 19. Leonenko, V.N., Ivanov, S.V.: Fitting the SEIR model of seasonal influenza outbreak to the incidence data for Russian cities. *Russian Journal of Numerical Analysis and Mathematical Modelling* **31**(5), 267–279 (2016)

20. Leonenko, V.N., Ivanov, S.V., Novoselova., Y.K.: A computational approach to investigate patterns of acute respiratory illness dynamics in the regions with distinct seasonal climate transitions. *Procedia Computer Science* **80**, 2402–2412 (2016)
21. Leonenko, V.N., Ivanov, S.V.: Influenza peaks prediction in Russian cities: comparing the accuracy of two SEIR models. *Mathematical Biosciences and Engineering* **15**(1), 209–232 (2018). <https://doi.org/10.3934/mbe.2018009>
22. Leonenko, V.N., Novoselova, Y.K., Ong, K.M.: Influenza outbreaks forecasting in Russian cities: Is Baroyan-Rvachev approach still applicable? *Procedia Computer Science* **101**, 282–291 (2016)
23. Leonenko, V.N., Pertsev, N.V., Artzrouni, M.: Using high performance algorithms for the hybrid simulation of disease dynamics on CPU and GPU. *Procedia Computer Science* **51**, 150–159 (2015)
24. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. *Mathematical programming* **45**(1-3), 503–528 (1989)
25. van Noort, S.P., Águas, R., Ballesteros, S., Gomes, M.G.M.: The role of weather on the relation between influenza and influenza-like illness. *Journal of theoretical biology* **298**, 131–137 (2012)
26. Public Health Dynamics Lab: FRED wiki. [online], <https://github.com/PublicHealthDynamicsLab/FRED/wiki/>
27. Seleznev, N.E., Leonenko, V.N.: Absolute humidity anomalies and the influenza onsets in Russia: a computational study. *Procedia computer science* **119**, 224–233 (2017)
28. Seleznev, N.E., Leonenko, V.N.: Boosting performance of influenza outbreak prediction framework. In: *Digital Transformation and Global Society*. pp. 374–384. Springer International Publishing (2017). <https://doi.org/10.1007/978-3-319-69784>
29. Shaman, J., Pitzer, V.E., Viboud, C., Grenfell, B.T., Lipsitch, M.: Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol* **8**(2), e1000316 (2010)
30. Tamerius, J., Nelson, M.I., Zhou, S.Z., Viboud, C., Miller, M.A., Alonso, W.J.: Global influenza seasonality: reconciling patterns across temperate and tropical regions. *Environmental health perspectives* **119**(4), 439 (2011)
31. UN: 68% of the world population projected to live in urban areas by 2050. [online], <https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html>
32. Wheaton, W.D., Cajka, J.C., Chasteen, B.M., Wagener, D.K., Cooley, P.C., Ganapathi, L., Roberts, D.J., Allpress, J.L.: Synthesized population databases: A US geospatial database for agent-based models. *Methods report (RTI Press)* **2009**(10), 905 (2009)
33. WHO: Influenza (seasonal). Fact sheet No. 211, March 2014. [online], <http://www.who.int/mediacentre/factsheets/fs211/en/>
34. Wilson, E.B., Worcester, J.: The law of mass action in epidemiology. *Proceedings of the National Academy of Sciences* **31**(1), 24–34 (1945)
35. Yandex: Auditorii. [online], <https://audience.yandex.ru/>
36. Yandex: Geocoder. [online], <https://tech.yandex.com/maps/geocoder/>