

# Contextual Contact Tracing based on Stochastic Compartment Modeling and Spatial Risk Assessment

Mateen Mahmood (✉ [al394323@uji.es](mailto:al394323@uji.es))

Universitat Jaume I <https://orcid.org/0000-0002-7784-3706>

Jorge Mateu

Universitat Jaume I

Enrique Hernández Orallo

Polytechnic University of Valencia: Universitat Politecnica de Valencia

---

## Research Article

**Keywords:** Compartment Modeling, Contact Tracing, Digital Epidemiology, Human Mobility, Self Organizing Maps, Trajectories

**Posted Date:** April 30th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-455884/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Stochastic Environmental Research and Risk Assessment on October 26th, 2021. See the published version at <https://doi.org/10.1007/s00477-021-02065-2>.

# Contextual Contact Tracing based on Stochastic Compartment Modeling and Spatial Risk Assessment

Mateen Mahmood · Jorge Mateu ·  
Enrique Hernández-Orallo

Received: date / Accepted: date

**Abstract** The current situation of COVID-19 highlights the paramount importance of infectious disease surveillance, which necessitates early monitoring for effective response. Policymakers are interested in data insights identifying high-risk areas as well as individuals to be quarantined, especially as the public gets back to their normal routine. We investigate both requirements by the implementation of disease outbreak modeling and exploring its induced dynamic spatial risk in form of risk assessment, along with its real-time integration back into the disease model. This paper implements a contact tracing-based stochastic compartment model as a baseline, to further modify the existing setup to include the spatial risk. This modification of each individual-level contact's intensity to be dependent on its spatial location has been termed as *Contextual Contact Tracing*. The results highlight that the inclusion of spatial context tends to send more individuals into quarantine which reduces the overall spread of infection. With a simulated example of an induced spatial high-risk, it is highlighted that the new spatio-SIR model can act as a tool to empower the analyst with a capability to explore disease dynamics from a spatial perspective. We conclude that the proposed spatio-SIR tool can be of great help for policymakers to know the consequences of their decision prior to their implementation.

**Keywords** Compartment Modeling · Contact Tracing · Digital Epidemiology · Human Mobility · Self Organizing Maps · Trajectories

---

M. Mahmood  
Universitat Jaume I, Castellón, Spain  
E-mail: al394323@uji.es

J. Mateu  
Universitat Jaume I, Castellón, Spain  
E-mail: mateu@uji.es

E. Hernández-Orallo  
Universitat Politècnica de Valencia, Valencia, Spain  
E-mail: ehernandez@disca.upv.es

## 1 Introduction

Detection and control of COVID-19 in particular, and infectious diseases in general, have irrupted as a major societal challenge. As of 31<sup>st</sup> January 2021, the COVID-19 pandemic has over 101 million confirmed cases with above 2.1 million deaths worldwide (WHO, 2021). This explosive dissemination is not only a universal threat to public health organizations, but it also jeopardizes social functioning, industry, economy and international relations (Zhou et al, 2020). Countries such as Israel and South Korea which took prompt actions towards testing and identification of previous contacts in case of an identified individual were able to restrict the disease spread. However, countries that did not proceed with the initial massive testing and contact tracing had to go for extreme measures of lockdown, quarantining and contact precautions (social-distancing, facemasks, etc.) (Hernández-Orallo et al, 2020).

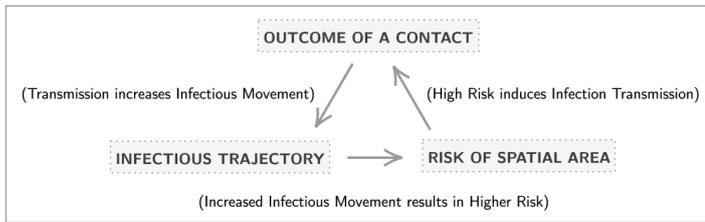
Detecting all infected individuals among the population requires massive testing on a regional scale. Though authorities have followed ingenious medical methods to rapidly detect the infected individuals, it has a considerable economical burden and implementation barriers. In a situation like this, detection of an infectious disease requires non-pharmaceutical interventions (NPI) and is to be supported by methods outside of the medical system, which sets the basis of the term *Digital Epidemiology* (DE) (Salathé, 2018).

One such DE based method is *Digital Contact Tracing* (DCT), which can provide prior contacts of a detected individual. This rapid identification of exposed individuals (who need to be tested or quarantined), can support the health system by restricting the uncontrolled asymptomatic propagation of infection. In DCT, the key to track the infectious transmission is to keep an eye on the physical *interaction* (contacts) of individuals, and understanding these interactions are as important as understanding the contagion process.

These interactions are much more than just recording of a *contact*, and when studied from a spatio-temporal perspective, they provide a comprehensive understanding of disease dynamics. While the temporal domain deals with the duration and instance of contacts, the spatial aspect refers the influence of a geographical location on the outcome of a contact, with a notion that some areas are inducing disease transmission more than others due to their urban function (Wang et al, 2017), environment and overall infectious activities.

On the other hand, these interactions based on individuals movement is subject to tracking of human mobility, where detection of an infected individual means that infectious trajectories can be tracked. Such tracking is critical to understand how an infection propagates in population and in space, as it not only identifies future infectious contacts but also highlights the places these infectious trajectories have visited (Benreguia et al, 2020). Identification of such high-risk areas is critical for policymakers in decisions related to smart lockdown, areal curfew, etc.

This scenario makes contact tracing, mobility tracking and spatial risk interconnected processes. It is a recursive sequence as illustrated in Figure 1, where the probability of transmission of a contact is proportional to the



**Fig. 1** Relation of temporally varying spatial risk affecting epidemic model and vice versa

73 risk intensity of its spatial location. This spatial location evolves based on  
 74 infectious movement, which itself is an outcome of an infectious contact. Hence,  
 75 there is a requirement of an approach to thoroughly fuse the effect of space  
 76 into a disease model while dealing with infectious trajectories. In this paper,  
 77 we focus on the inclusion of the spatial aspect of these physical interactions  
 78 termed as *Contextual Contact Tracing*. The idea is that contacts taking place  
 79 in contextually distinct geographical locations are to be treated differently  
 80 based on the vulnerability they pose to the susceptible individual.

81 In human infectious diseases, where the pathogen is another human being,  
 82 there is a requirement to track human movement. Tracking the known infected  
 83 individuals and their interactions are already demanding, but the existence  
 84 of asymptomatic individuals makes this monitoring even more challenging  
 85 (Müller et al, 2020). These undetected individuals are transmitting the infection  
 86 to a larger set of individuals, who themselves are infecting the community  
 87 in an uncontrollable domino effect. Early detection of asymptomatic individ-  
 88 uals followed by isolation or treatment is the key to restrict pandemic growth,  
 89 where state of the art highlights the accepted practise of digital methods in  
 90 such detection studies (Anglemyer et al, 2020). Ongoing research (Van Dore-  
 91 malen et al, 2020; Simmerman et al, 2010) highlights the aerosol and surface  
 92 stability of infectious diseases, where COVID-19, SARS-CoV1 and Influenza  
 93 A/H1N1, all have indicated up to days surface transmission. Both these as-  
 94 pects, tracking of individuals and risk assessment of space, sets the basis of  
 95 infectious disease surveillance in this digital era.

96 Tracking human movement relies on human mobility data, which is of  
 97 prime importance in individual-level research on infectious disease dynamics  
 98 (Brockmann et al, 2009). Recent advancement in location-aware technologies  
 99 and computing procedures have resulted in a massive influx of mobility data,  
 100 which is capable of representing the movement of an individual to a very small  
 101 scale up to less than a meter (Zheng, 2015). This high-level detail makes these  
 102 datasets an ideal candidate for high precision tasks such as contact tracing.

103 Despite that, an important consideration is that continuous recording of  
 104 an individual movement is highly invasive (Reichert et al, 2020), which is  
 105 why there is no infectious disease related individual-level trajectory dataset  
 106 publicly available so far. To minimize this concern, the use of bluetooth has  
 107 been proposed (Martinez-Martin et al, 2020), though it only collects the con-

108 tact information as and when it happens. At the same time, Benreguia et al.  
109 (Benreguia et al, 2020) suggest that in preparation for an extremely critical  
110 scenario where entire humanity is at stake and the requirement of saving lives  
111 is of highest priority, the use of continuous recordings of individual’s movement  
112 is justified given it is implemented by government and with a guarantee of pri-  
113 vacy protection. Similarly, for spatial risk assessment, individual-level work is  
114 only executed on a sparse scale. Souza et al. (Souza et al, 2019) detected spa-  
115 tial clusters using spatial scan statistics, based on Twitter feed data. Another  
116 spatial clustering application on aggregated data is available in (Desjardins  
117 et al, 2020) where a countywide space-time clustering is executed.

118 The long-standing COVID-19 has amplified research in this domain with  
119 several studies involving individual-level-mobility for investigation of disease  
120 dynamics. Many of these studies described the spatio-temporal trends inclusive  
121 of stochastic aspects, proposing statistical foundations to fit models to data.  
122 However, the spatial aspects focused more on spatial separation rather than  
123 spatial location. Even if the spatial location was considered, it was in the  
124 aggregated form of spatially varying demographic factor (Mahsin et al, 2020).

125 In epidemic modeling, compartment models distribute each individual in  
126 the population based on their disease states. Generally, they are of Susceptible,  
127 Infected and Recovered (SIR) type, however many versions such as SEIR,  
128 SEIAHCRD (Berger et al, 2020; Bardina et al, 2020) exist which depend on  
129 the type of disease and applied methodologies. Though the temporal aspect is  
130 well addressed in these SIR models, the *spatial context* is generally new.

131 In spatio-epidemic modeling, the idea of a space-dependent SIR model has  
132 been presented in (Takács and Hadjimichael, 2019) in form of a numerical  
133 experiment. They considered a generalized SIR model where population size  
134 differed over space. Another spatial-SIR model is explained in (Bisin and Moro,  
135 2020) to understand spatial diffusion of disease based on quantitative effects  
136 of geographical context in determining that diffusion. Modifying epidemic pa-  
137 rameters based on the spatial location have also been proposed. A space-time  
138 dependent *basic reproductive ratio* is implemented in (Martinez-Beneito et al,  
139 2020), while Lang et al. (Lang et al, 2018) discuss a framework of a SIR model  
140 on spatial networks where the probability of transmission is based on spatial  
141 distances along the edges. A *bayesian maximum entropy* based extension is also  
142 available for metapopulation-level epidemic modeling (Angulo et al, 2013).

143 All these models propose population-level frameworks for the inclusion  
144 of space in SIR modeling. Complete integration of a spatial context in an  
145 individual-level study of contact tracing is still missing, which can consider  
146 the influence of *space* (location) for each specific contact.

147 This paper proposes a new spatio enhanced setup of SIR modeling, where  
148 contacts are associated with an intensity of its risk score based on its spatial  
149 location. This association of risk with a contact is executed by reforming the  
150 quantitative value of a contact, where enhancement is in a manner that a riskier  
151 contact has a higher probability of disease transmission than the one which  
152 is of relatively lower risk. For temporally varying spatial risk, we re-evaluate  
153 spatial risk scores based on infectious activities of the recent past.

154 Here, we analyze real-life mobility data of NCCU Trace (Tsai and Chan,  
155 2015) which provides movement of 115 students recorded for 15 days. In the im-  
156 plementation, we first execute contact tracing to construct temporal network  
157 graphs. These contact graphs are further used to implement an epidemic model  
158 with self-induced infection, which was later enhanced to a spatially-enhanced  
159 epidemic model including the spatial risk. In parallel, we track infectious tra-  
160 jectories and the location of contacts as elements for spatial risk assessment.  
161 The results highlight that the inclusion of spatial context tends to send more  
162 individuals into quarantine which reduces the overall spread of infection.

163 The reason behind pursuing this study in the absence of real information  
164 about infection is because a methodology that considers spatial risk in a con-  
165 tact tracing process is also missing. Therefore, the feasibility of this idea is  
166 developed in the form of a *spatio-epidemic tool*, which is an established pro-  
167 posal for future works, not only to work with real datasets as they become  
168 available but also in the domain of spatial risk.

169 A recent publication from February 2021 presents movement data of in-  
170 fected (COVID-19) individuals from South Korea (Park et al, 2021). However,  
171 data is not in the form of continuous trajectories, but are recordings of indi-  
172 vidual's interactions with others through a contact tracing application. This  
173 availability is a motivating fact as more real-world datasets related to infection  
174 information as well as mobility trajectories will be publicly available offering  
175 a definite way forward for this work.

176 The remainder of this paper is as follows: Section 2 describes the method-  
177 ology of both, SIR model (baseline setup) and its enhancement to a spatio-SIR  
178 model. Section 3 introduces the selected dataset with discussion on the exper-  
179 imental design. This section further presents results of both models supported  
180 by varying simulations to effectively understand the new spatio-SIR setup.  
181 Section 4 concludes the paper and presents limitations and future work.

## 182 2 Methodology

### 183 2.1 Baseline-SIR modeling

184 The baseline-SIR model for this study is motivated by (Hernández-Orallo et al,  
185 2020), in which contact tracing technologies are evaluated along with a compar-  
186 ison of stochastic versus deterministic approaches. In this paper, we reproduce  
187 their stochastic setup (hereby referred to as base-SIR) as our baseline-SIR  
188 model, with a rationale that a stochastic model is more realistic than a deter-  
189 ministic one due to its probabilistic nature. Similarly, an event-based method  
190 as followed in base-SIR is preferred due to the incorporation of event-driven  
191 chance element. An overview of the methodology for the implementation of  
192 baseline-SIR, and its modification to a spatio-SIR model (see Section 2.2) is  
193 available in Figure 2.

194 Base-SIR brings forth a novel addition of *Quarantine Susceptible* and *Quar-*  
195 *antine Infected* related compartments which add a new perspective in the mod-

196 eling of a real-world scenario related to contact tracing based compartment  
 197 modeling. Base-SIR implements *Gillespie’s First Reaction Method* (GFRM)  
 198 (Keeling and Rohani, 2011), which handles efficiently a contact tracing prob-  
 199 lem, especially on a trajectory-based dataset.

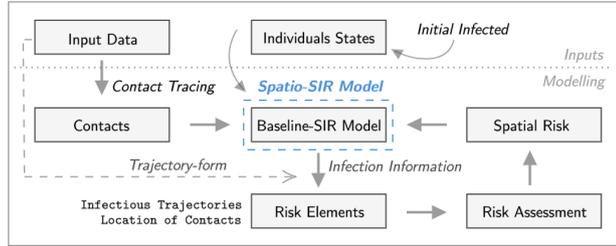


Fig. 2 Summarized workflow of overall methodology

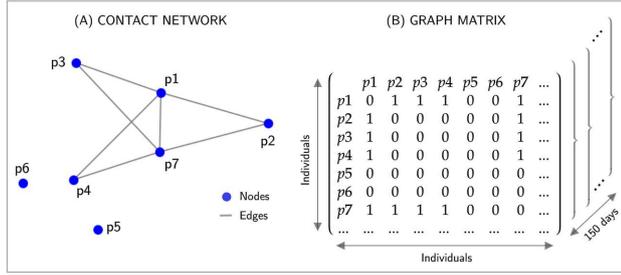
### 2.1.1 Contact Tracing

201 Contact Tracing is the identification of *colocation* of two or more individuals.  
 202 However, this colocation is not restricted to a single point or a single instance  
 203 of time, but a range of an area and duration which are based on epidemiological  
 204 aspects. Contact with a possibility of transmission is the one within two meters  
 205 of an infected individual with an exposure of at least one minute (Hernández-  
 206 Orallo et al, 2020). Therefore, we define  $d_c$  as the *distance threshold* and  $t_c$  as  
 207 the *duration threshold* for considering a contact as risky.

208 Identification of infectious contacts needs accurate information about the  
 209 possible transmissible pathways from an infected person to each individual in  
 210 the population (Eames and Keeling, 2003). A network graph is a computa-  
 211 tionally efficient representation of such interactions where in individual-level  
 212 studies, *nodes* refer to individuals and *edges* represent their contacts (Enright  
 213 and Kao, 2018). A temporal network graph can be denoted as  $G(t)$ , with  $\nu$   
 214 (nodes) and  $\varepsilon$  (edges), where  $t$  represents the instance of time. In epidemic  
 215 modeling, it is common to have a temporal frequency of a *day* (Keeling and  
 216 Rohani, 2011), hence  $\varepsilon_{ij}(t)$  will exist between individual  $i$  and  $j$  if there exists  
 217 a contact between the two on day  $t$ .

218 An adjacency matrix is commonly used to store graph information. It is a  
 219 *graph matrix*, where rows and columns represent nodes (individuals), with a  
 220 third dimension corresponding to the day of contact. A contact is represented  
 221 with a value of either 0 or 1, where 1 depicts the existence of an edge (contact)  
 222 between the two. Figure 3 presents a toy example of a network graph and the  
 223 associated graph matrix.

224 Directions of edges as in directed/undirected graph are ignored as contacts  
 225 are independent of direction. This highlights the assumption that infection  
 226 can be transmitted in both directions depending on the disease state of the



**Fig. 3** A toy example: contacts of a single day in the form of contact network and graph matrix

227 individual and not the structure of the network. For a pair  $(i, j)$  of individuals,  
 228 this symmetry can be viewed as  $(G_{ij}(t) = G_{ji}(t))$ .

229 In contact network, *degree* shows the count of connections of a node with  
 230 the other nodes in the network. The *temporal degree*  $K_i(t)$  is the count of  
 231 contacts of a person  $i$  with other individuals in the network  $G(t)$  on day  $t$ .  
 232 Hence, an *average degree*  $\kappa$  for a time period  $T$  can be computed as (1)

$$233 \quad \kappa = \frac{1}{N} \sum_{i=1}^N \left[ \frac{1}{T} \sum_{j=1}^T K_i(t_j) \right] \quad (1)$$

234 As the rate of infection is influenced by the count of infected individuals,  
 235 hence it is useful to have a degree only involving contacts with infected indi-  
 236 viduals. Such a degree of diffusion can be represented as (2), where  $I_j(t)$  is an  
 237 indicator function denoting that individual  $j$  can infect others,

$$238 \quad \mathcal{K}_i(t) = \sum_{j=1}^N G_{ij}(t) \cdot I_j(t) \quad (2)$$

239 Identifying prior contacts is the overall essence of contact tracing in order  
 240 to restrict next generation of cases. This requires a backward time window  
 241  $\Delta$  depending on the type of disease (infectious period, incubation time, etc.),  
 242 and can be used in the form of (3) to extract all prior contacts  $\mathcal{C}_i(t, \Delta)$  of an  
 243 individual  $i$  at  $t$  with window  $\Delta$

$$244 \quad \mathcal{C}_i(t, \Delta) = \sum_{j=1}^N \left( \max_{\tau \in [t-\Delta, t]} G_{ij}(\tau) \right) D_j(t) \quad (3)$$

245 Here,  $D_j(t)$  is 1 if at time  $t$  person  $j$  is infected and traced. Algorithm (1)  
 246 explains the process flow of contact tracing. Once the contacts are identified,  
 247 a baseline setup can be formulated to simulate SIR events. The model also  
 248 evaluates the efficiency of contact tracing methods. Contact tracing can be  
 249 manual (that is, based on interviewing the detected and infected individuals)  
 250 or smartphone based (using contact tracing apps). We define a value  $q$  as

251 the fraction of traced individuals being quarantined. For example, this value  
 252 can reflect the number of individuals that use the mobile contact tracing app.  
 253 In the case where the tracing time is greater than 1, the  $q$  value must be  
 254 normalised by the average tracing time ( $1/\tau_T$ ), as  $q' = q/(1/\tau_T) = q \cdot \tau_T$   
 255 in order to distribute the tracing quarantine over the days. The idea is that  
 256 if the tracing time is long (for example, by using interviews), it is precisely  
 257 because it takes time to trace back the prior contacts, so the whole number  
 258 of traced individuals during this tracing time is equally distributed over these  
 259 days. Finally, apart from contacts, baseline-SIR model relies on (a) *infection*  
 260 *states* and (b) *epidemic parameters*.

---

**Algorithm 1** *Pseudocode of contact tracing process*


---

```

input: trace_dataset, n                                ▷ ( $n \leftarrow$  count of individuals)
output: G                                             ▷ ( $G \leftarrow$  contact graph)
// initialize  $G(n, n, d) \leftarrow 0$                     ▷ ( $d \leftarrow$  number of days)
1: for ( $t \leftarrow 1$  to  $d \times 86400$ ) do                 ▷ ( $t \leftarrow$  time in seconds)
2:   for ( $p1 \leftarrow 1$  to  $n$ ) do                         ▷ ( $p1 \leftarrow$  first person)
3:     for ( $p2 \leftarrow 2$  to  $n$ ) do                       ▷ ( $p2 \leftarrow$  second person)
4:       if ( $p1 \neq p2$ ) then
           dist = distance between p1 and p2
5:         if ( $\text{dist} < d_c$ ) then                          ▷ ( $d_c \leftarrow$  distance threshold)
6:           record duration
7:           if ( $\text{duration} > t_c$ ) then                    ▷ ( $t_c \leftarrow$  duration threshold)
8:              $G(p1, p2, \text{day}) \leftarrow 1$                 ▷ ( $\text{day} \leftarrow t/86400$ )
9: return: G

```

---

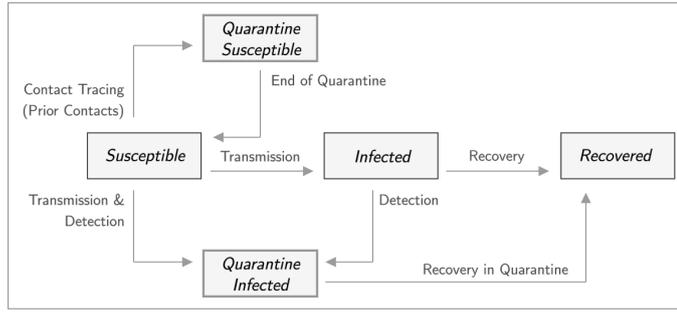
261 

### 2.1.2 Infection states (compartments)

262 Infection states refer to the compartments an individual can be during an  
 263 epidemic. As in base-SIR, a total of five compartments are considered which  
 264 represent the states of Susceptible ( $S$ ), Infected ( $I$ ), Recovered ( $R$ ), Quarantine  
 265 Susceptible ( $Q_S$ ) and Quarantine Infected ( $Q_I$ ). With these five compart-  
 266 ments, there are seven possible SIR events that will imply the transition of  
 267 an individual from a compartment to another. Figure 4 presents the possible  
 268 events (transfers among compartments), which are:

269 ( $S \rightarrow I$ ,  $S \rightarrow Q_S$ ,  $S \rightarrow Q_I$ ,  $I \rightarrow Q_I$ ,  $Q_S \rightarrow S$ ,  $I \rightarrow R$ ,  $Q_I \rightarrow R$ )

270 As information about the latency states of individuals is not available, a  
 271 self-induced infection approach is followed. This means that out of the total  
 272 population, a certain count of individuals in the population are initiated as  
 273 infected being in compartment ( $I$ ), to have a sense of disease propagation  
 274 based on their future contacts, as epidemic progresses.



**Fig. 4** Overview of compartments and possible transfers between them

### 2.1.3 Epidemic parameters

Epidemic parameters, as introduced in Table 1, refer to the disease-specific elements in the form of coefficients that contribute to computing the rates of each event associated with individuals.

**Table 1** Summary of infectious disease related modeling parameters

parameter	description
$\kappa$	Average degree (daily contacts per individual)
$\kappa_i$	Contacts of individual $i$ with infected individuals
$\mathcal{R}_0$	Basic reproductive ratio
$\delta$	Rate of detection
$\gamma$	Recovery rate ( $1/\gamma =$ disease specific days for recovery)
$b$	Probability of transmission of infection ( $b = \mathcal{R}_0 \cdot \gamma/\kappa$ )
$\beta$	Transmission rate ( $\beta = \kappa \cdot b$ )
$\tau_Q$	Time in quarantine
$q'$	Tracing efficiency ( $q' = q \cdot \tau_T$ )
$C_i(\Delta)$	Backward contact tracing of individual $i$ with detected infected individuals

The core of the model is to answer the question of how individuals move from one compartment to another. In a closed environment where births, deaths and migration are ignored, transition ( $S \rightarrow I$ ) is subject to disease transmission and is a function of three aspects: (i) the presence of infected individuals, (ii) contacts between susceptible and infected ( $S \leftrightarrow I$ ) and (iii) the probability of transmission. Considering  $\kappa$  as the *degree* of ( $S \leftrightarrow I$ ) contacts and  $b$  representing the *probability of transmission* of infection, the *transmission rate*  $\beta$  can be written as  $\beta = \kappa \cdot b$ . The transition from ( $I \rightarrow R$ ) is simpler as it can be considered a constant around a mean value based on clinical data of *infectious period*. The probability of an infected individual to be recovered relies on how long they have been infected, which can be denoted as *recovery rate*  $\gamma$ , a constant value representing the inverse of infectious period. The ratio

**Table 2** Rate equations related to each SIR event

event	description of rate	rate equation
$S \rightarrow I$	Transmission of infection	$(1 - C_i(t, \Delta)) \cdot b \cdot \mathcal{K}_i(t)$
$S \rightarrow Q_S$	Susceptible person being quarantined	$q' \cdot C_i(t, \Delta) \cdot (1 - b \cdot \mathcal{K}_i(t))$
$S \rightarrow Q_I$	Susceptible person being infected and detected	$q' \cdot C_i(t, \Delta) \cdot b \cdot \mathcal{K}_i(t)$
$I \rightarrow Q_I$	Infected person being detected	$\delta$
$Q_S \rightarrow S$	End of quarantine after quarantine period	$\mathcal{T}_Q$
$I \rightarrow R$	Recovery after infectious period	$\gamma$
$Q_I \rightarrow R$	Recovery after infectious period in quarantine	$\mathcal{T}_Q$

291  $\beta/\gamma$  is called *basic reproductive ratio*  $\mathcal{R}_0$ . It represents the expected count of  
 292 cases directly affected by a single case and is considered as one of the more  
 293 representative parameters of disease in epidemiology.

294 Similarly, based on the epidemic parameters enlisted in Table 1, rates of  
 295 each event can be computed using equations provided in Table 2. In this paper,  
 296 we propose a generalized framework for spatio-SIR modeling through the use of  
 297 values corresponding to COVID-19, as given in Table 3; however, any disease-  
 298 specific model can be developed by adjusting these parameters.

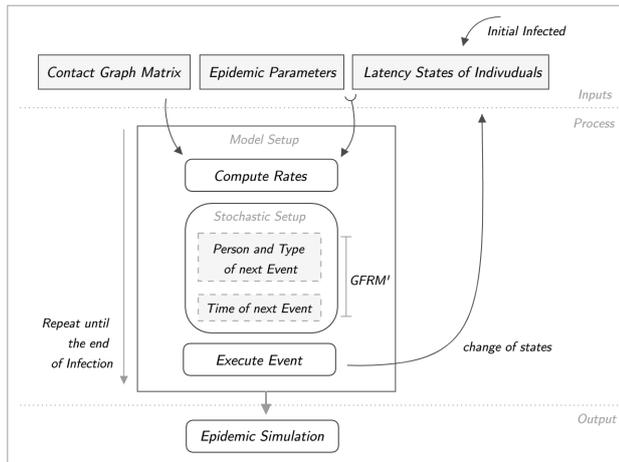
**Table 3** Estimated values for COVID-19 (extracted from (Ferretti et al, 2020; Hellewell et al, 2020; Li et al, 2020))

parameter	estimated value
$\mathcal{R}_0$	3
$\delta$	0.1
$\gamma$	1/15
$\mathcal{T}_Q$	1/14

#### 299 2.1.4 Event-based modeling

300 In an event-driven model, each possibility is considered as an event and then  
 301 a random element will decide which event may happen, based on the cumu-  
 302 lative rates of all events and converting those rates into probabilities. This  
 303 highlights that even if the probability of an event is similar, an individual  
 304 may experience a varied event based on a random or stochastic effect. There  
 305 are numerous methods to implement event-driven approaches, one of which is  
 306 *Gillespie's Method* (Gillespie, 1977), common in SIR modeling (Keeling and  
 307 Rohani, 2011).

308 Gillespie's algorithm, initially intended for the study of chemical reactions,  
 309 is also applicable in scenarios such as SIR modeling where an outcome of the  
 310 contact is like a biochemical reaction of a cell with fluctuating possibilities  
 311 of events. It is a variant of a Monte Carlo method, with a computationally



**Fig. 5** Process flow of an event-driven stochastic SIR model

312 feasible solution. Gillespie’s First Reaction Method (GFRM) is a simplified  
 313 version of the original Gillespie’s Direct Method with a scalable approach.

314 There are two stochastic elements in GFRM. The first one is the *type*  
 315 of event which includes the person over which the event will happen, and  
 316 the kind of event (out of the defined seven events) that will happen. The  
 317 second stochastic element is the *time* of the next event, which refers to the  
 318 duration since the previous event. The former, as per GFRM, is determined  
 319 by computing the rates of each event and then stochastically drawing the next  
 320 event. The latter, in our approach, is completely stochastic based on a random  
 321 element instead of computing the time for each event. This modification is  
 322 due to the fact that there is no inherent time of an event in a contact tracing  
 323 process. Based on this modified GFRM, event-based stochastic SIR model  
 324 can be implemented on identified contacts using *infection states* and *epidemic*  
 325 *parameters*. Figure 5 shows the workflow of GFRM-based SIR model.

## 326 2.2 Spatio-SIR modeling

327 This section focuses on two aspects: *temporally varying spatial risk* and *spatio-*  
 328 *SIR model*. Here, spatial risk refers to the transmission vulnerability a spatial  
 329 location poses to a susceptible individual involved in an infectious contact.  
 330 This spatial risk is for a certain period and is continuously evolving based on  
 331 previous infectious activities. Our spatio-SIR model extends the baseline setup  
 332 taking into account the spatial risk in the future tracing of contacts.

333 As the goal is to associate a risk score to each contact based on its spatial  
 334 location, it is important to address the definition of *location*. A simple and  
 335 computationally efficient approach is to consider a regular lattice (grid) struc-  
 336 ture segmenting the study area into smaller cells, each one having a risk score.

337 From this, location of a contact can be defined as the corresponding cell of the  
 338 grid in which the contact is taking place.

### 339 2.2.1 Temporally varying spatial risk

340 In this study, the spatial risk relies only on monitoring of SIR events, to track  
 341 infectious trajectories and location of contacts (Benreguia et al, 2020). With  
 342 such monitoring, we computed risk scores based on four risk basis, as follows:

343 (a) *Infectious trajectories* refer to the amount of time an infectious trajec-  
 344 tory has spent in each cell, where an infected person is only tracked from  
 345 the time of infection until recovery or quarantine.

346 (b) *Infected individuals* refer to the count of infected individuals in each  
 347 cell. This is distinct from (a) in the sense that shorter cumulative duration  
 348 of many individuals is riskier than a longer duration of a single individual.

349 (c) *Infectious contacts* refer to the locations of all contacts involving an  
 350 infected individual. They are more than the count of times infection is ac-  
 351 tually transmitted as this involves all ( $S \leftrightarrow I$ ) contacts; and on the other  
 352 end, transmission is dependent on the randomly chosen event. From this,  
 353 a relation *transmissive contacts*  $\in$  *all* ( $S \leftrightarrow I$ ) *contacts*, can be deduced.

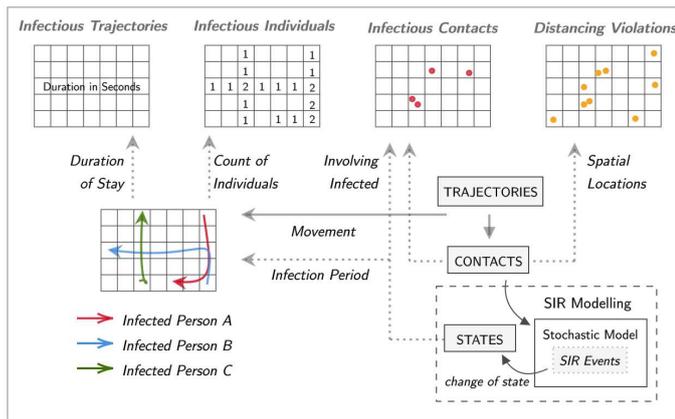
354 (d) *Social distancing violations* refer to all ( $S \leftrightarrow S$ ) contacts. This property  
 355 reflects population density and also captures the notion that a place (cell)  
 356 with higher precautionary violations must be of higher risk than a place  
 357 following the public health regulations (Rezaei and Azarmi, 2020).

358 As contact graphs are developed per day, the same frequency can be fol-  
 359 lowed in order to develop these four risk types. This means that risk scores  
 360 of each cell are based on the cumulative effect of activities from the previous  
 361 day and are to be updated every next day. Figure 6 depicts the process of  
 362 computing risk basis, by tracking infectious trajectories for their duration and  
 363 count, alongside the monitoring of contacts for their spatial locations.

364 With four risk basis, there comes a need for integrating these risk attributes  
 365 into a single representation. This requires normalizing all grids to a common  
 366 range and to further combine them into a single grid. This results in a *risk*  
 367 *map*, based on activities from the previous day to provide an evolved risk for  
 368 the next day.

### 369 2.2.2 Unsupervised classification using SOM and K-means

370 For combining multiple aspects into a single map form, implementation of a  
 371 multi-criteria analysis approach is not applicable as there is no prior informa-  
 372 tion of which criterion is significant over others. For classification, a supervised  
 373 method requires information about the characteristics of the target class and  
 374 pre-existing labels for the method to cluster data and label them accordingly.  
 375 However, lack of validation data restricts the application of supervised classifi-  
 376 cation. A possible solution is to implement an unsupervised learning method,  
 377 as it does not rely on pre-existing labels for reinforcement. Such methods only



**Fig. 6** Computing risk basis from trajectories and contacts

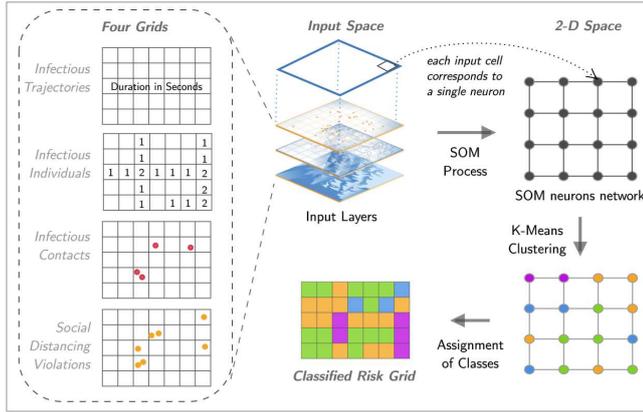
378 require input patterns to highlight relationships and can assist in the explo-  
 379 ration of the available covariates to develop a single classified risk map.

380 One such unsupervised clustering technique is *Self Organizing Map* (SOM),  
 381 which can serve to the purpose of combining information of multiple grids  
 382 into a single one. SOM is basically a dimensionality reduction technique but  
 383 as SOM preserves the topographic relationships in feature space to ensure  
 384 nearby objects are clustered together, it has been extensively used for the  
 385 clustering of geospatial data (Henriques et al, 2012; Gopal, 2016).

386 Considering the dimensionality reducing capability, SOM is similar to the  
 387 statistical equivalent of Principal Component Analysis, whereas Baccao et  
 388 al. (Bação et al, 2005) suggest SOM as a possible substitute for K-means  
 389 clustering when the neighbourhood is not considered. Besides, in comparison  
 390 to statistical techniques, SOM offers three main advantages due to its non-  
 391 parametric nature: (i) it works independent of variable's distributions, (ii) it  
 392 is computationally efficient to non-linear problems and (iii) it caters for noise  
 393 or missing data more effectively (Asan and Ercan, 2012).

394 As highlighted in (Vesanto and Alhoniemi, 2000), the best approach to  
 395 implement SOM is a two-step process. First, input data is to be transformed  
 396 into a two-dimensional neurons network; secondly, SOM neurons are to be  
 397 clustered using a hierarchical or partitive approach. The major benefits of this  
 398 two-step approach are: (i) the computational efficiency even with a smaller  
 399 dataset and (ii) noise reduction in case of imperfect data as input for clustering.

400 An important consideration here is to choose the size of SOM neurons net-  
 401 work which is dependent on the size of input dataset. In our implementation,  
 402 we use a regular lattice of  $12 \times 15 = 180$  cells, hence an optimal size of SOM  
 403 neurons can be acquired as  $5 \cdot \sqrt{180} = 67.08$ . For a two-dimensional structure  
 404 of neurons network, we considered a total of 64 neurons instead of 67, which  
 405 could be arranged in a symmetric shape of 8x8 neurons network.



**Fig. 7** Unsupervised classification workflow using SOM and K-means

406 After the establishment of SOM network as a representation of input data  
 407 (multiple grids), a hierarchical clustering process can cluster the neurons into  
 408 the desired number of groups. In this paper, we follow the partitioning ap-  
 409 proach of K-means clustering as they do not rely on previously found clusters  
 410 as the hierarchical approach does (Vesanto and Alhoniemi, 2000).

411 As risk values of grid cells vary over time, and a flexible value of  $k$  (number  
 412 of clusters) in K-means can result in a different count of classes for different  
 413 days, we fixed the count of classes to 5 so as to have an equal number of  
 414 groups every time a new spatial risk is computed. However, in case data does  
 415 not allow to have five classes, then an optimized number of classes is chosen  
 416 for an appropriate representation. This results in a classified grid-based risk  
 417 where each cell corresponds to a class of risk. As the output after K-means is an  
 418 un-ordered classification, which is the same as segmenting the cells in different  
 419 groups but not knowing which group is of higher risk, classes are assigned with  
 420 appropriate labels by comparing the cumulative average of the risk score in  
 421 all cells associated with each class and further assignment of ordered labels  
 422 in descending order with greater average as the highest risk class. Figure 7  
 423 illustrates the complete process of combining multiple grids through the use  
 424 of SOM followed by K-means and further labelling.

### 425 2.2.3 Contextual contact tracing

426 In order to include the temporally varying spatial risk for each specific contact,  
 427 we modify the previous contact graph  $G$  to obtain a new contact graph  $G'$ .  
 428 For this modification, we use the daily-acquired risk-based grids (spatial risk)  
 429 and based on the location of the contacts we obtain  $G'$  which considers the  
 430 risk score of each contact's location. In  $G'$ , each contact value has a varying  
 431 intensity depending on the spatial risk compared to the constant value of 1  
 432 (which represented a contact) of the base-SIR model. Here, we introduce a

433 new range representation for a contact between the value of 0.5 (*lowest risk*)  
 434 to 1.5 (*highest risk*). The rationale behind this range is to be able to compare  
 435 with baseline-SIR setup (see Figure 13) where the previous value 1 is the mean  
 436 of the new range representation. Once real data about spatial risk as well as  
 437 infection information are available to fit the model to the data, different values  
 438 for this range can be configured to identify the best fit.

439 Using this new matrix  $G'$  in (2), we obtain a new *degree of diffusion*  $\mathcal{K}'_i(t)$ ,  
 440 which is used on the rate equations defined on Table 2. As rates of events in  
 441 SIR model are based on the cumulative infectious contacts represented by  $\mathcal{K}_i$ ,  
 442 a varying contact value (between 0.5 to 1.5) will result in a varied influence  
 443 to the transmission process for each specific contact, meaning a direct effect  
 444 of spatial risk on the disease transmission. As in the baseline-SIR model, this  
 445 spatio-SIR model can be solved using the GFRM's rate equations as stated  
 446 in Table 2. Besides, the consideration of varying  $\mathcal{K}'_i$  in these equations only  
 447 influences events related to *Susceptible* population ( $\mathbf{S} \rightarrow \mathbf{I}$ ,  $\mathbf{S} \rightarrow Q_S$  and  $\mathbf{S} \rightarrow$   
 448  $Q_I$ ). However, there is no influence of spatial risk on events related to *Infected*  
 449 individuals and those in *Quarantine*. This process of dynamically computing  
 450 risk scores based on daily movement and reflecting its effect by modifying  
 451 contact graph is termed as *Contextual Contact Tracing*.

### 452 3 Data analysis and simulations

#### 453 3.1 Dataset and experimental design

454 NCCU Trace (Tsai and Chan, 2015) refers to an android application to trace  
 455 movements of 115 students in a campus environment of National Chengchi  
 456 University (Taiwan), for a period of 15 days with measurement interval up  
 457 to 10 minutes and spatial position rounded to meters. The application was  
 458 designed to capture information regarding GPS, WiFi, and Bluetooth devices  
 459 in proximity, resulting in their movement traces. The Appendix contains de-  
 460 tails of the NCCU dataset with an overview of the study area and sample of  
 461 recordings. For an epidemic, a period of 15 days is very short to assess the  
 462 spread of infection. A possible solution is to extend the period of the dataset  
 463 by concatenating the same dataset multiple times, as the pattern of human  
 464 mobility shows a regularity over the same weekdays. Such joining can produce  
 465 a data for 150 days, an appropriate duration for epidemic modeling.

466 Both baseline-SIR and the spatio-SIR models are evaluated over NCCU  
 467 data. The experiments assume 10 individuals as *initial infected* ( $I_0 = 10$ ) on  
 468 the first day of the epidemic with no recovered individual ( $R = 0$ ). Sum of indi-  
 469 viduals in all compartments is 115 at all times. We used the values of COVID-  
 470 19 parameters as discussed in Table 3 and *tracing efficiency*  $q'$  of 0.1. For the  
 471 stochasticity, 10 realizations of the same initial conditions but the random al-  
 472 location of initial infection are executed. This means that in each realization,  
 473 infected individuals are different. Averaging the results over 10 realizations,  
 474 average curves are obtained, where a curve represents the count of individuals

475 in each compartment. Due to stochasticity, duration of the epidemic in these  
 476 realizations varies, hence we extrapolate trends of other realizations to the  
 477 epidemic with the longest duration to obtain an average representation.

478 In each run within a single realization of a model, only one epidemic event  
 479 is executed. The *time* of next event is a stochastic duration as a part of day,  
 480 hence there are multiple events per day, with at least one event in a day, and  
 481 overall hundreds of events even for a short epidemic of few weeks. Hence, using  
 482 GFRM, a single realization of the model moves forward executing events after  
 483 each *time* step. As a whole, these executed events simulate a disease outbreak  
 484 scenario.

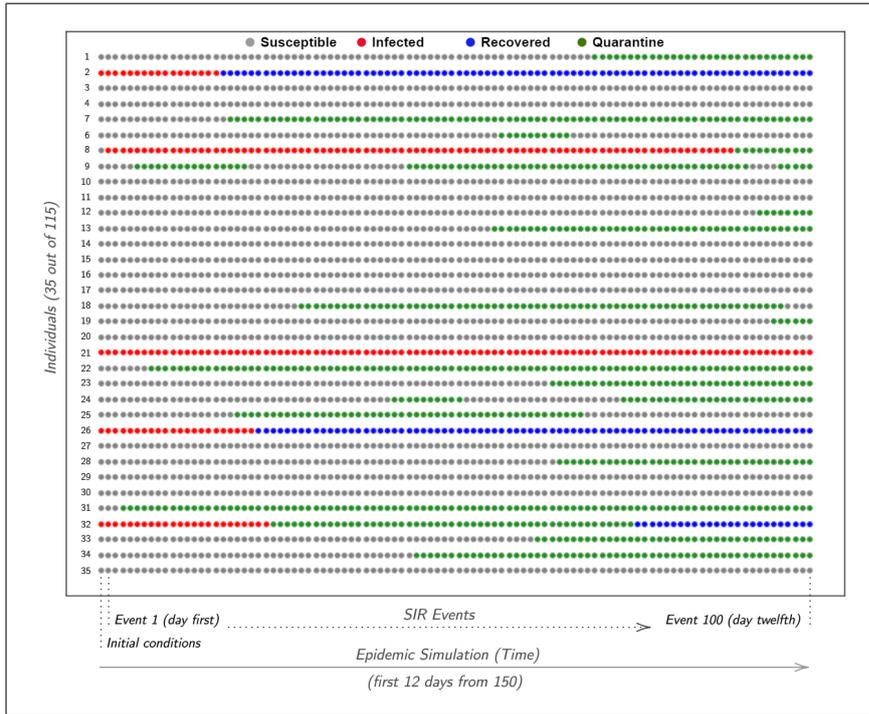
## 485 3.2 Data analysis

### 486 3.2.1 Results - Baseline-SIR model

487 The ability of an individual-level compartment model to monitor the latent  
 488 state of each individual at all times highlights its importance in the infectious  
 489 diseases realm. To understand this capability, Figure 8 illustrates individual-  
 490 level latency of a subset of the population (35 out of 115). At the start of  
 491 the epidemic (day 1), four individuals (2, 21, 26 and 32) are infected as the  
 492 initial outbreak, whereas the remaining all are *Susceptible*. The first stochastic  
 493 event (second column from left) is of infection for individual (8). In every  
 494 iteration, there is only one event, where the time of the next event (a part  
 495 of the day) is also random, hence there can be multiple events in a single  
 496 day. Individual (8) remains infected for a week and gets detected around the  
 497 11<sup>th</sup> day. Individual (2) gains recovery only after few days. Individual (21)  
 498 remains infected and undetected for the whole shown period. Similarly, the  
 499 state of each individual can be observed based on the time-series review of  
 500 their associated compartment.

501 Exploring the modification of a spatial context needs the setup of a base-  
 502 line model to experiment over. Figure 9 presents the output of such a baseline  
 503 setup in form of an outbreak scenario using parameters from Table 1. At the  
 504 beginning of the epidemic, everyone except the *Infected* is in the *Suscepti-*  
 505 *ble* compartment, which means there is no *Recovered* individual. Initially, the  
 506 count of *Infected* individuals increases from 10 to 14 in the first few days as  
 507 *Susceptible* population interacts (contacts) with already infected (initial out-  
 508 break). However, not only their count decreases afterwards as they are sent  
 509 into *Quarantine Infected*, but the *Susceptible* count also diminishes from initial  
 510 count of 105 to 40 in a fortnight. Due to backward tracing  $\mathcal{C}_i$ , a higher number  
 511 of individuals are identified as *exposed* and sent into *Quarantine Susceptible* as  
 512 a precautionary measure. These plummeted trends of the count of *Susceptible*  
 513 and *Infected* forces less population on the streets, which not only restricts the  
 514 future infectious contacts but ultimately the overall disease outbreak.

515 The peak of individuals in *Quarantine Susceptible* is around the 19<sup>th</sup> day  
 516 with 40 plus individuals, where afterwards the sum of individuals remains

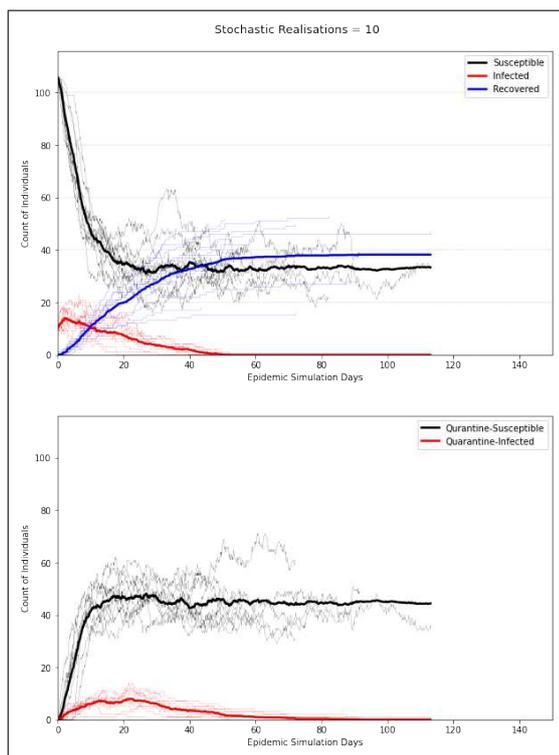


**Fig. 8** Individual-level change in latency of 35 out of 115 total individuals is shown based on the SIR events as model runs forward. Each row belongs to a single individual, where the compartment they belong to at an instance of time is represented column-wise chronologically from left to right. There is only one event per column with multiple events per day, where figure illustrates first 100 events from the initial 12 days of epidemic.

517 nearly constant which depicts an equal frequency of individuals moving be-  
 518 tween ( $S \leftrightarrow Q_S$ ) compartments. *Quarantine Infected* compartment reaches its  
 519 highest count of 5 twice on the 13<sup>th</sup> and 22<sup>nd</sup> days. Once a person is *Recov-*  
 520 *ered*, that individual remains in that compartment, which is evident from the  
 521 continuous increase in its count from 0 at the start to 38 at its end. Even  
 522 after there is no *Infected* person on the street after the 45<sup>th</sup> day, the model  
 523 continues in anticipation of risk due to the presence of individuals in *Quar-*  
 524 *antine Infected*; and ultimately ends the epidemic with their recovery around  
 525 the 113<sup>th</sup> day.

526 *3.2.2 Results - Enhanced spatio-SIR model*

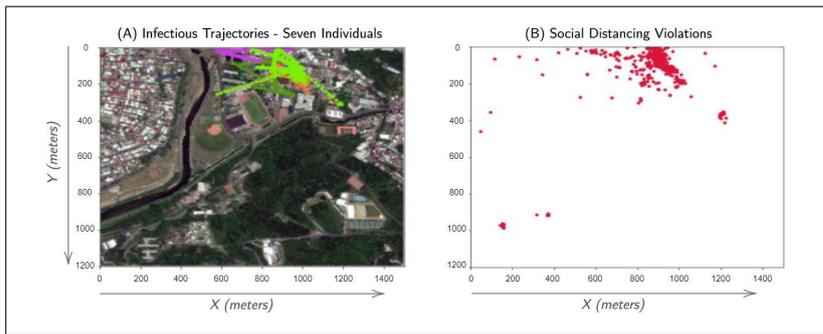
527 Spatio-SIR enhancement is achieved by computing the spatial risk out of events  
 528 in the baseline setup, which requires monitoring of SIR events for infectious  
 529 activities. Figure 10 presents a 1-day sample of such infectious activities. Figure  
 530 10-A illustrates the movement of infectious individuals shown over the study  
 531 area. In this sample, there are 7 infected individuals with mobility concentrated



**Fig. 9** Output of a disease outbreak scenario by baseline-SIR model. (*Top*) presents trends related to counts of *Susceptible*, *Infected* and *Recovered*, whereas (*Bottom*) illustrates count of individuals in *Quarantine* related compartments. Each of the 10 realizations of stochastic model is shown (*light in color*), with their average curves represented with (*dark bold*) lines. Count of total population is 115 which are represented over the *Y-axis*.

532 inside the NCCU campus (center-top). Out of these infectious trajectories, two  
 533 sorts of attributes are extracted. First is the collective duration of time spent  
 534 by these individuals in each area, and secondly how many individuals were  
 535 located in each area. The other two basis are of *Infectious Contacts* and *All*  
 536 *Contacts*, where the latter is shown in Figure 10-B. It identifies locations of  
 537 all contacts termed as *social distancing violations* in order to highlight the  
 538 notion that a place with a higher number of contacts means it is of higher  
 539 risk than a place with a lower number of contacts. This concept has been also  
 540 implemented by (Rezaei and Azarmi, 2020) for infection risk assessment.

541 Based on the risk basis shown in Figure 10, grid-based risks are developed  
 542 as presented in Figure 11. Here, the trajectories and contacts are transformed  
 543 into a grid structure with an intensity of associated attributes normalized to  
 544  $[0,1]$ . Figures 11-A and 11-B capture information of infectious trajectories in  
 545 terms of *duration* and *count* respectively. Similarly, the location of different  
 546 nature of contacts is captured in Figures 11-C and 11-D. Based on the previous  
 547 day, these attributes serve as the basis of risk for the next day.

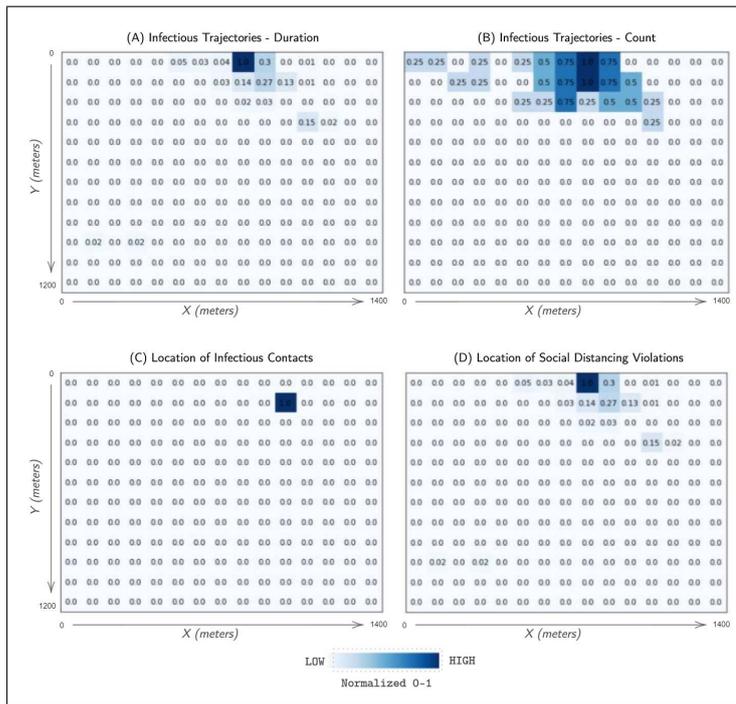


**Fig. 10** Risk basis of *infectious trajectories* and *location of contacts* from one complete day during an epidemic. In the sample shown in Sub-Figure (A), there are total of 7 infected individuals from that day. Sub-Figure (B) shows locations of all contacts from the same day.

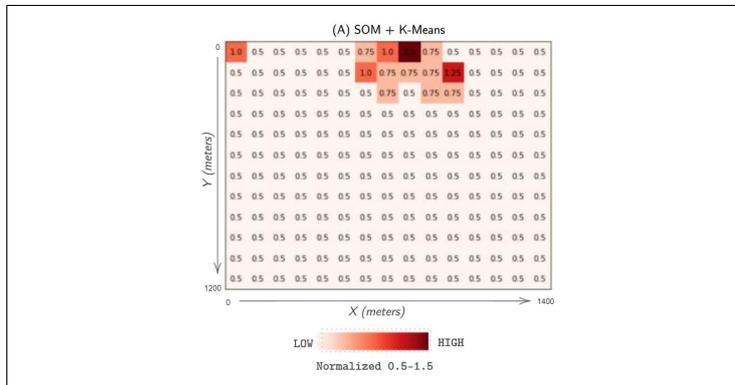
548 To identify spatial risk for the future contextual tracing of contacts, multi-  
 549 grids from Figure 11 are integrated into a single grid as shown in Figure  
 550 12. In order to classify the output to segment areas of higher or lower risk,  
 551 risk scores are grouped into 5 classes with their labels corresponding to their  
 552 intensity of risk. The classes of risk are (0.50, 0.75, 1.00, 1.25, 1.50) with 1.50  
 553 referring to the highest risk. A review of this result shows that based on activi-  
 554 ties from the previous day (Figure 10), the highest risk area is at the centre-top  
 555 cells, whereas the surrounding areas are also of higher risk. While there is no  
 556 spatial risk in the remaining study area on this particular day, however, due to  
 557 the temporally varying nature, the spatial risk may evolve in future instants.

558 Results of the spatio-SIR model are compared with the results of baseline-  
 559 SIR in Figure 13. As the inclusion of spatial risk tends to affect the rates  
 560 of events related to *Susceptible* individuals and getting infected is subject to  
 561 an infectious contact, hence in the spatio-SIR model, there are more events  
 562 of the population moving into *Quarantine Susceptible*. Though the trends of  
 563 *Quarantine Susceptible* in both models are similar till day 15<sup>th</sup>, however, the  
 564 mentioned phenomenon is evident afterwards where the peak of individuals in  
 565 *Quarantine Susceptible (spatio-SIR)* is 59 on the 29<sup>th</sup> day, whereas there are  
 566 less than 50 individuals in *Quarantine Susceptible (baseline-SIR)* by the same  
 567 day.

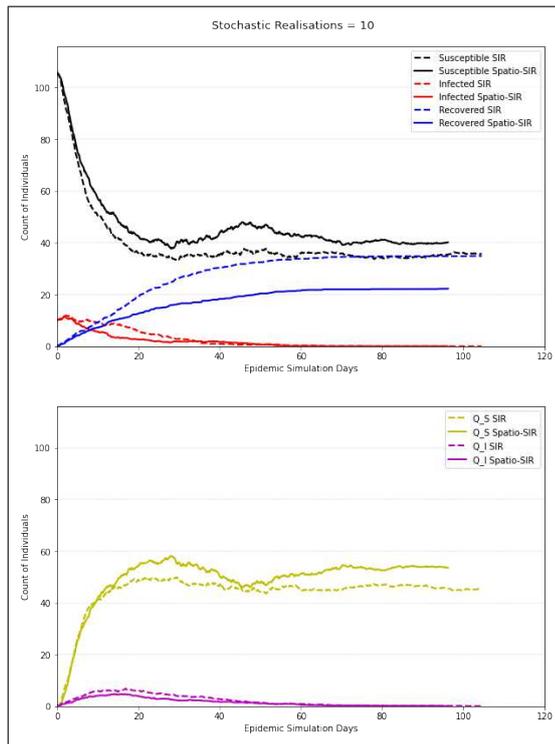
568 Early events of quarantining reduces the counts of *Infected*, *Recovered* and  
 569 *Quarantine Infected*. Comparing the trends of *Susceptible* population, it can  
 570 be observed that in the first week both are more or less similar, however,  
 571 the first week onward the susceptible population in baseline-SIR decreases to  
 572 40 by the 17<sup>th</sup> day, whereas it takes an extra week (23<sup>rd</sup> day) for the same  
 573 decline up to 40 in spatio-SIR. This highlights that due to the additional  
 574 aspect of spatial risk, a greater fraction of the population remains susceptible.  
 575 Similarly, an increase in the count of *Susceptible* around the 45<sup>th</sup> day depicts  
 576 the return of quarantined population after a period of two weeks, whereas such  
 577 a return is not visible in baseline-SIR as there is no consideration of spatial



**Fig. 11** Grid-based risk computed out of risk basis. Sub-Figure (A) and (B) refer to grid-based representation of *duration* and *count* of trajectories shown in Figure 10-(A). Sub-Figure (C) depicts location of an infectious contact occurred on the same day, whereas Sub-Figure (D) presents translation of Figure 10-(B) into grid form. All values are normalized to the range of 0 to 1, where 1 refers to highest risk.



**Fig. 12** Combining risk from multiple grids shown in Figure 11 into a single grid output using SOM and K-means. This integration is executed in unsupervised manner through the implementation of SOM followed by K-means. Risk scores are computed in the range of [0.5,1.5], where 1 refers to the previous normal (existing SIR model with a constant spatial risk and all contacts being of equal nature)

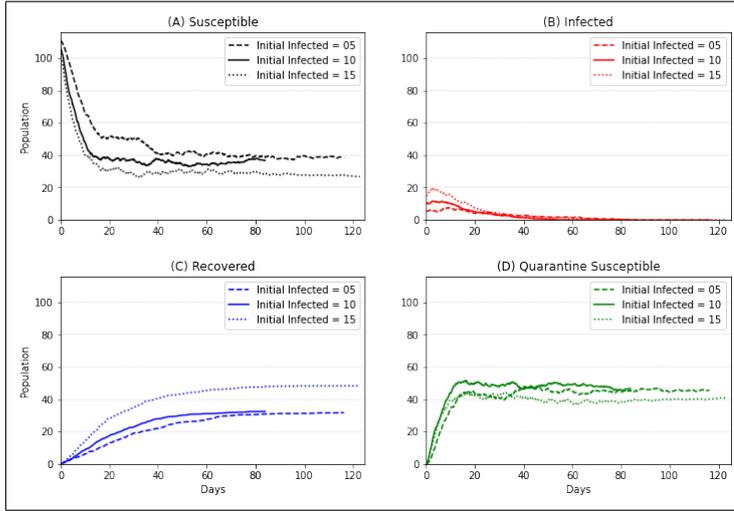


**Fig. 13** Comparing *average* of 10 stochastic realizations of a disease outbreak scenario from baseline-SIR (*dashed*) and spatio-SIR (*solid*). (*Top*) presents trends related to count of *Susceptible*, *Infected* and *Recovered*, whereas (*Bottom*) illustrates counts of individuals in *Quarantine* related compartments. Count of total population is 115 which are represented over *Y-axis*.

578 risk. With a higher count of total individuals in *Quarantine Susceptible*, the  
 579 overall infection is controlled which can be confirmed from the trend of *Infected*  
 580 and *Recovered*. As in spatio-SIR model, the total recovered are 21 compared  
 581 to the count of 35 in baseline-SIR model. The same can be observed in the  
 582 trends of *Quarantine Infected*, as with less *Infected* on the streets, the spread  
 583 of infection is controlled; hence, a lower count of *Quarantine Infected* in spatio-  
 584 SIR compared to baseline-SIR, apart from the start and end of an epidemic  
 585 which is nearly similar.

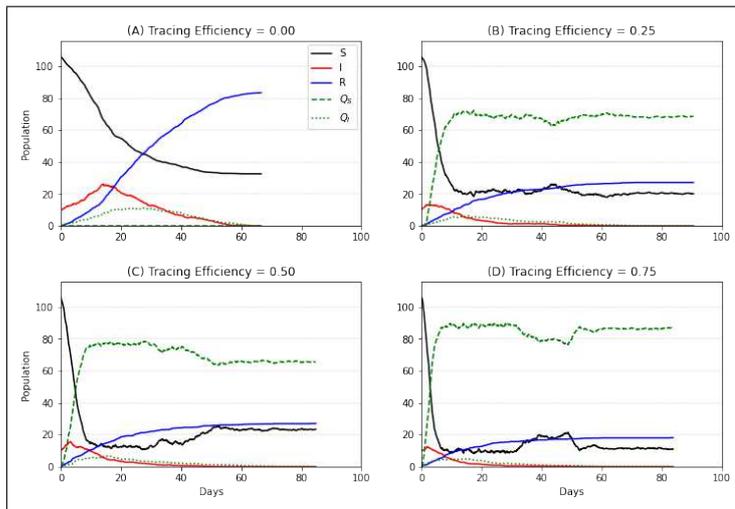
### 586 3.2.3 Simulated scenarios over the NCCU data

587 This section reinforces the need for a spatio-epidemic tool. As the model simu-  
 588 lates a scenario based on the initial values, changing the initial setup can help  
 589 assess impact of the change on the overall disease outbreak simulation. Here,  
 590 baseline-SIR model executes one such variation as presented in Figure 14 with  
 591 different intensities of the *Initial Infected*  $I_0$ .



**Fig. 14** Comparing the average of 10 stochastic realizations (of baseline-SIR model) with varying count of *Initial Infected*  $I_0$ , to observe their effect on the overall disease outbreak. Three scenarios of  $I_0 = 5/10/15$  are shown with a population size of 115. Subplots (A, B, C and D) shows comparison of *Susceptible*, *Infected*, *Recovered* and *Quarantine Susceptible*, respectively. Due to stochasticity, final duration of epidemic varies depending on the overall spread of infection.

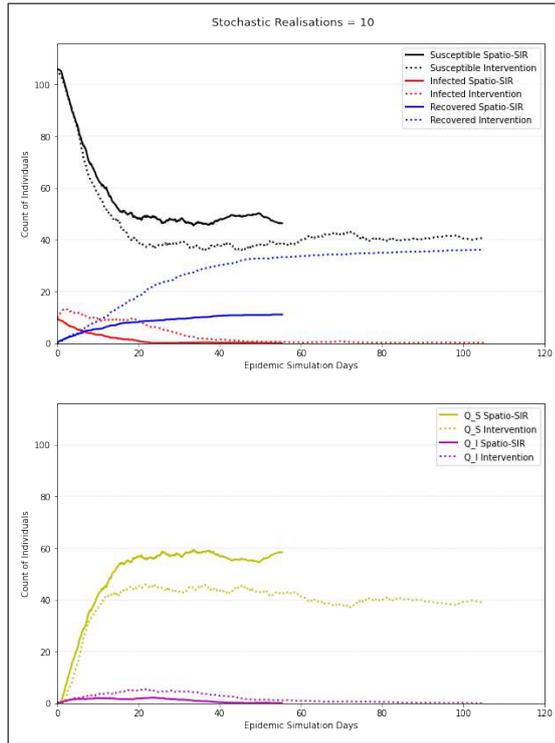
592 In general, the higher quantum of initial outbreak results in a longer epi-  
 593 demic which is evident in all subplots. In Figure 14-A, *Susceptible* population  
 594 is compared, where higher count of initial outbreak reflects in early departure  
 595 of individuals from the susceptible compartment; either getting *Infected* (due  
 596 to greater frequency of infectious contacts) or *Quarantined* (because of prior  
 597 contact tracing of *Infected* individuals). Higher infected counts ( $I_0 = 10$  and  
 598  $I_0 = 15$ ) result in decrease of *Susceptible* count from 105/100 to approximately  
 599 40 within 2 weeks, whereas when  $I_0 = 5$  reaches the count of 40 after six weeks.  
 600 Figure 14-B illustrates the effect of varying initial outbreak on the total counts  
 601 of *Infected*, where a directly proportional relationship is evident in the initial  
 602 spread of infection up to the 19<sup>th</sup> day. However, once a majority of *Infected* are  
 603 sent into *Quarantine Infected* and a higher count of individuals are already in  
 604 *Quarantine Susceptible*, all scenarios tend to have a similar pattern afterwards.  
 605 Similarly, Figure 14-C depicts a likewise trend of initial difference, where two  
 606 setups of ( $I_0 = 05$ ) and ( $I_0 = 10$ ) later (after 70<sup>th</sup> day) coincide to have a  
 607 similar pattern (around 30 *Recovered* individuals). However, ( $I_0 = 15$ ) results  
 608 in a massive outbreak with almost 50 *Infected* individuals by the 70<sup>th</sup> day.  
 609 Figure 14-D highlights that a higher count of initial infected will either send  
 610 more contacts into *Quarantine Infected* or *Quarantine Susceptible*, which is  
 611 dependent on (i) the *transmission rate* ( $\beta = \kappa \cdot b$ ) and (ii) the chance element  
 612 of event-based setup. Hence, the relation of initial infected with *Quarantine*  
 613 related compartments is not straightforward. However, the trend of ( $I_0 = 15$ )  
 614 specially after the 40<sup>th</sup> day depicts that due to greater initial outbreak, more



**Fig. 15** Comparing the average of 10 stochastic realizations (of baseline-SIR model) with varying *Tracing Efficiency*  $q'$ , to observe their effect on the overall disease outbreak (population size = 115). Subplots (A, B, C and D) present four cases of  $q' = 0, 0.25, 0.50$  and  $0.75$ , respectively, where 1 means 100% backward tracing.

615 individuals were *Infected*, thus more people are in *Recovered* and *Quarantine*  
 616 *Infected*, because of which the overall count of *Quarantine Susceptible* is low.

617 Another possible variation on the analyzed scenarios can be the *Tracing*  
 618 *Efficiency* which is available in Figure 15. Tracing efficiency refers to the frac-  
 619 tion of identified prior contacts based on backward tracing. As 100% tracing  
 620 is not plausible, only a proportion is evaluated as an estimate of tracing. In  
 621 the case of no backward tracing (zero efficiency) shown in Figure 15-A, there  
 622 are no individuals in *Quarantine Susceptible*. Only *Infected* who gets detected  
 623 are sent into quarantine, which results in a massive disease outbreak with a  
 624 count of *Recovered* more than 80 individuals. In Figure 15-B, 62 individuals  
 625 are in *Quarantine Susceptible* by the 10<sup>th</sup> day, whereas with efficiency of 0.50  
 626 (Figure 15-C) and 0.75 (Figure 15-D), there are 77 and 88 individuals in *Quar-*  
 627 *antine Susceptible* by the same period of 10 days. It can be deduced that for  
 628 every 25% increase in the *Tracing Efficiency*, 10% more population is sent into  
 629 quarantine. In general, it can be said that with greater tracing efficiency, the  
 630 greater amount of population is early forced for quarantine, which ultimately  
 631 reduces the overall spread of infection (less *Infected* and less *Recovered*). The  
 632 population being forced to quarantine means they are leaving the *Susceptible*  
 633 compartment, which is evident by the degree of slope in the downward trend  
 634 of *Susceptible* count proportional to tracing efficiency. Due to high tracing effi-  
 635 ciency in Figure 15-D, a huge subset of the population is sent into *Quarantine*  
 636 immediately as the infection breaks out. This large amount of individuals when  
 637 collectively comes out of quarantine (after a period of 14 days), results in a

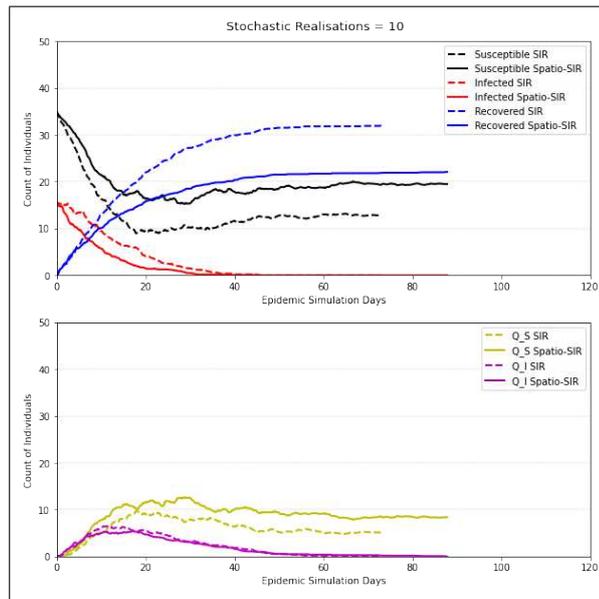


**Fig. 16** Comparing average of 10 stochastic realizations of a disease outbreak scenario from spatio-SIR (*solid*) and a case of Intervention - spatio-SIR model with spatial risk of 1.5 from day 11 to 20 (*dotted*). (*Top*) presents trends related to counts of *Susceptible*, *Infected* and *Recovered*, whereas (*Bottom*) illustrates counts of individuals in *Quarantine* related compartments. Count of total population is 115 which are represented over *Y-axis*.

638 sudden drop of  $Q_S$  count around the 40<sup>th</sup> day. An opposite can be observed  
 639 in the count of *Susceptible*.

640 Other than varying the initial configuration, another capability of our  
 641 spatio-SIR model is the ability to simulate real-world scenarios such as re-  
 642 laxation in social-distancing, spatio-temporal curfew/lockdown or a holiday  
 643 season with more population on the streets. This capability of the tool can  
 644 assist policymakers to simulate scenarios, visualizing the consequence of their  
 645 decisions prior to their actual implementation. One real scenario would be the  
 646 relaxation of distancing measures in the campus with an increase of in-class  
 647 teaching and in-campus social living, which will increase the people on the  
 648 campus and their mobility (and therefore, the risk of contact). In an experi-  
 649 ment presented in Figure 16, we introduce this scenario as an *Intervention*  
 650 in a specific period from day 11<sup>th</sup> to 20<sup>th</sup>. Quantitatively, this intervention is in  
 651 the form of spatial high-risk of value 1.5 at all areas (cells).

652 A major difference is in the overall period of epidemic, where the *Interven-*  
 653 *tion* setup executes an epidemic of more than 100 days considering the added



**Fig. 17** Comparing average of 10 stochastic realizations of a disease outbreak scenario from baseline-SIR (*dashed*) and spatio-SIR (*solid*). (*Top*) presents trends related to counts of *Susceptible*, *Infected* and *Recovered*, whereas (*Bottom*) illustrates counts of individuals in *Quarantine* related compartments. Count of total population is 50 which are represented over *Y-axis*.

654 spatial risk from day 11 to 20, whereas in spatio-SIR modeling the epidemic  
 655 finished in less than 60 days. Observing the trend of *Recovered* individuals, a  
 656 continuous increase after day 10 is evident in *Intervention* setup, compared to  
 657 spatio-SIR output. This escalation ends up with a total of 37 recovered in the  
 658 former, while total recovered individuals in the latter are 11. A similar pattern  
 659 is identifiable while observing the trends of *Infected* population, where since  
 660 day 10<sup>th</sup>, the rate of infection is more or less constant (a horizontal line) un-  
 661 til the 20<sup>th</sup> day. This is different from the infected trend in spatio-SIR model  
 662 where the rate of infection is decreasing after the initial increase in the first few  
 663 days of the epidemic. Observing the trend of *Quarantine Susceptible*, a spike  
 664 is noticeable after day 11 in the *Intervention* setup. Counts of susceptible in  
 665 quarantine in *Intervention* setup is 56 on day 20<sup>th</sup>, whereas under the spatio-  
 666 SIR model there are only 43 susceptible individuals in quarantine by the same  
 667 day, confirming the capability of new setup to capture spatial high-risk.

### 668 3.3 Mobility simulation in a new space

669 As individual-level mobility datasets are scarcely available, a possible solution  
 670 is to self-simulate movement trajectories for the study area (new space). This

671 can also help in the application of our methodology on multiple datasets to  
 672 assess its performance in different spaces.

673 In this study, we have also generated a synthetic dataset using spatial move-  
 674 ments from Geolife Data (Zheng et al, 2011). Geolife, a project by Microsoft,  
 675 provides trajectory movements of 178 users for a period of four years with  
 676 temporal resolution of 1 to 5 seconds and spatial resolution of 5 to 10 meters.  
 677 The dataset in total contains 17,621 trajectories, total distance of 1,251,654  
 678 kilometers holding information of 48,203 hours. As for contact tracing and  
 679 spatial risk assessment, we require mobility to be highly concentrated on a  
 680 small study area. Unfortunately, this is not the case in the original Geolife  
 681 dataset. However, we used only its spatial movement but modified the tem-  
 682 poral and user-related attribute to reflect the daily movement of 15 days for  
 683 50 users, for a study area of  $20 \times 16$  square kilometres. Figure 22 in the Ap-  
 684 pendix presents a visualization of this modified construction, whereas Figure  
 685 17 illustrates the comparison of the SIR model and spatio-SIR model over this  
 686 new dataset. Both models, SIR and spatio-SIR, depict a similar trend on syn-  
 687 thetic trajectories as over NCCU trace. The consideration of spatial risk tends  
 688 to send more people in *Quarantine Susceptible* which initially protects them  
 689 from the infection, but the population remains susceptible in general as the  
 690 quarantined population comes back to *Susceptible* stage after a quarantining  
 691 period. Similarly, the overall infection propagation is reduced due to a lower  
 692 count of infectious contacts which results in lower counts of *Recovered*.

#### 693 4 Conclusions and discussion

694 We conclude that the inclusion of spatial risk in epidemic modeling can greatly  
 695 support the public health system by identification of infectious contacts and  
 696 highlighting places carrying the high risk. It is a bi-fold domino effect that re-  
 697 lies on both, persons and places, and breaking the chain is necessary not only  
 698 in terms of individuals but also for high-risk areas. For a critical time such as  
 699 COVID-19, an integrated approach as the one introduced here can be devel-  
 700 oped into a comprehensive system of *infectious disease surveillance*. In terms  
 701 of modeling, consideration of spatial risk as in spatio-SIR model increases  
 702 the tracing efficiency, where a greater number of individuals are highlighted  
 703 as exposed depending on the location of contacts; as in this study contacts  
 704 are mostly concentrated in a small region that is at high-risk at all times.  
 705 These vulnerable individuals who are currently *Susceptible* will either be in-  
 706 fected or sent into quarantine, depending on the chance element of event-based  
 707 setup. This consideration of exposure based on spatial risk tends to perform  
 708 more *meaningful events*<sup>1</sup> concerned with the *Susceptible* population rather  
 709 than events to the *Infected* or *Quarantined*. Furthermore, it is shown that this  
 710 framework can act as a tool for policymakers to execute scenarios, visualizing  
 711 the consequence of their decisions prior to their actual implementation.

---

<sup>1</sup> Infection or Quarantining, compared to Recovery while in Quarantine

712 We have proposed a generalized framework for spatio-SIR modeling, how-  
713 ever, a disease-specific model can be developed by adjusting the parameters  
714 available in Table 3. With regards to contact tracing, the study highlights  
715 that for contact tracing to be effective, the maximum fraction of the popu-  
716 lation needs to be digitally activated, using the contact tracing app or other  
717 implemented mode of tracking (Hernández-Orallo et al, 2020).

718 The major limitation of the study is the non-availability of actual informa-  
719 tion about infected individuals. With that, the proposed methodology could  
720 have been configured to fit a model to data. In this paper, this limitation  
721 was handled through a self-induced initial outbreak. Another aspect is that  
722 the selected dataset is not of an epidemic scenario. A dataset from an era of  
723 an epidemic situation can assist in the analysis of such patterns and further  
724 explore its spatial risk. Similarly, 15 days recording of movements is a non-  
725 adequate period for a long-standing scenario of an epidemic. In this paper,  
726 this limitation was handled by concatenating the same dataset multiple times  
727 for 150 days. However, a better option would be to have a mobility dataset  
728 of a longer duration. A limitation of the followed approach is that the con-  
729 tacts were identified per day. This approach helped in establishing a setup to  
730 understand disease dynamics in a spatial context, however, a finer frequency  
731 such as hourly contact graphs or a real-time application of tracing in terms of  
732 recording a contact as they happen can be followed for higher accuracy.

733 This modification of an existing SIR model into a spatio-SIR model through  
734 the inclusion of spatial risk serves only as a foundation of an idea. This leads to  
735 many way forwards opening new avenues for the integration of spatial compo-  
736 nent into digital epidemiology. Spatial risk is a complete domain in itself that  
737 includes the identification of factors stimulating the vulnerability of being in-  
738 fected at a certain place and time. Hence, it is recommended to incorporate the  
739 spatial context from additional perspectives other than just infectious trajec-  
740 tories. A suggested idea is to integrate spatial information such as points of in-  
741 terests (restaurants, parks, etc.), public transits, urban functions, demographic  
742 details and environmental factors, for the overall spatial risk assessment. Such  
743 a study will explore the spatial effect of covariates in disease transmission by  
744 understanding their intrinsic underlying relationships presenting a higher or  
745 lower score of risk, and additionally, how these covariates amalgam as a whole.

746 Implementation of this study was based on an event-based SIR model where  
747 rates of events were computed to randomly draw the next event, as well as  
748 the time of the event and the person to which event will occur. This complete  
749 stochasticity can be adjusted in a sense to develop a semi-stochastic setup  
750 where the person to which event will occur is not completely random but a  
751 factor based on their exposure. Such a factor can be associated with each  
752 individual based on their movement in infectious places and the frequency of  
753 their contacts. Though semi-stochastic in nature, a specific model like this can  
754 also provide with the exposure profile for each individual.

755 Given that this domain of infectious diseases generally lacks data avail-  
756 ability related to infection and/or movement, a practical way forward is to  
757 transform this spatio-enhanced model into a comprehensive tool for simula-

758 tions. Such a tool can allow users to feed in movement data and then based  
759 on infectious movements, the user can execute spatio-SIR modeling while con-  
760 figuring the initial setup. Furthermore, the tool can have the capabilities to  
761 implement real-world scenarios like spatial curfew, commercial lockdown, re-  
762 laxation in social distancing, etc. The overall situation of COVID-19 signifies  
763 the importance of such a tool that can support public health policymakers as  
764 and when required.

765 Overall, this paper concludes that tracking of individual-level infectious  
766 trajectories is critical not only for person-to-person contact tracing but also  
767 to identify spatial risk which is transmitting (surface/aerosol transmission)  
768 as well as propagating (inducing riskier contact) in nature. The study also  
769 highlights that accurate modeling of this sort is restricted due to the data un-  
770 availability, and there is a critical requirement of datasets to ensure a practical  
771 application of the proposed approach.

772 The authors conclude this study with the remarks, that even if this domain<sup>2</sup>  
773 is generally hindered due to the lack of data availability, the investigation pro-  
774 cess related to it should keep on exploring methods to effectively understand  
775 disease dynamics. This is beneficial not only for literature but also critical for  
776 the overall well being of humanity.

---

<sup>2</sup> individual-level trajectory-based infectious diseases SIR modeling

## References

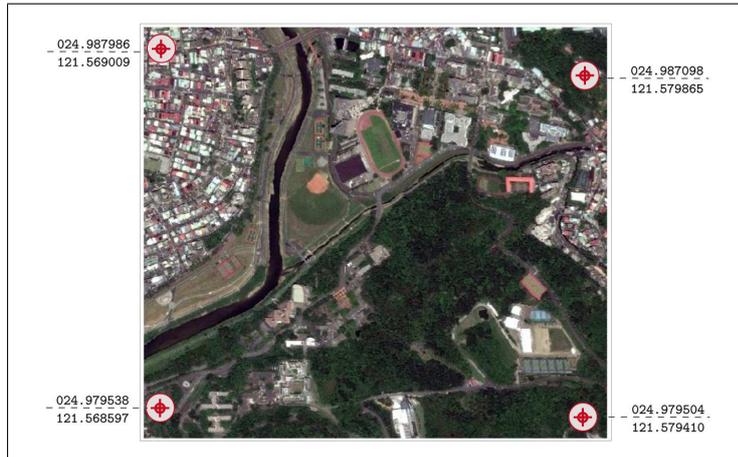
- 778 Anglemeyer A, Moore TH, Parker L, Chambers T, Grady A, Chiu K, Parry M,  
779 Wilczynska M, Flemyng E, Bero L (2020) Digital contact tracing technolo-  
780 gies in epidemics: a rapid review. *Cochrane Database of Systematic Reviews*  
781 (8) [3](#)
- 782 Angulo J, Yu HL, Langousis A, Kolovos A, Wang J, Madrid AE, Christakos  
783 G (2013) Spatiotemporal infectious disease modeling: a bme-sir approach.  
784 *PloS one* 8(9):e72,168 [4](#)
- 785 Asan U, Ercan S (2012) An introduction to self-organizing maps. In: *Computa-*  
786 *tional Intelligence Systems in Industrial Engineering*, Springer, pp 295–315  
787 [13](#)
- 788 Bação F, Lobo V, Painho M (2005) Self-organizing maps as substitutes for  
789 k-means clustering. In: *International Conference on Computational Science*,  
790 Springer, pp 476–483 [13](#)
- 791 Bardina X, Ferrante M, Rovira C (2020) A stochastic epidemic model of covid-  
792 19 disease. *arXiv preprint arXiv:200502859* [4](#)
- 793 Benreguia B, Moumen H, Merzoug MA (2020) Tracking covid-19 by tracking  
794 infectious trajectories. *arXiv preprint arXiv:200505523* [2, 4, 12](#)
- 795 Berger DW, Herkenhoff KF, Mongey S (2020) An seir infectious disease model  
796 with testing and conditional quarantine. *Tech. rep.*, National Bureau of Eco-  
797 nomic Research [4](#)
- 798 Bisin A, Moro A (2020) Learning epidemiology by doing: The empirical impli-  
799 cations of a spatial-sir model with behavioral responses. *Tech. rep.*, National  
800 Bureau of Economic Research [4](#)
- 801 Brockmann D, David V, Gallardo AM (2009) Human mobility and spatial  
802 disease dynamics. *Reviews of nonlinear dynamics and complexity* 2:1–24 [3](#)
- 803 Desjardins M, Hohl A, Delmelle E (2020) Rapid surveillance of covid-19 in the  
804 united states using a prospective space-time scan statistic: Detecting and  
805 evaluating emerging clusters. *Applied Geography* p 102202 [4](#)
- 806 Eames KT, Keeling MJ (2003) Contact tracing and disease control. *Pro-*  
807 *ceedings of the Royal Society of London Series B: Biological Sciences*  
808 270(1533):2565–2571 [6](#)
- 809 Enright J, Kao RR (2018) Epidemics on dynamic networks. *Epidemics* 24:88–  
810 97 [6](#)
- 811 Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dorner L, Parker  
812 M, Bonsall D, Fraser C (2020) Quantifying sars-cov-2 transmission suggests  
813 epidemic control with digital contact tracing. *Science* [10](#)
- 814 Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions.  
815 *The journal of physical chemistry* 81(25):2340–2361 [10](#)
- 816 Gopal S (2016) Artificial neural networks in geospatial analysis. *International*  
817 *Encyclopedia of Geography: People, the Earth, Environment and Technol-*  
818 *ogy: People, the Earth, Environment and Technology* pp 1–7 [13](#)
- 819 Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Munday  
820 JD, Kucharski AJ, Edmunds WJ, Sun F, Flasche S, Quilty BJ, Davies N,  
821 Liu Y, Clifford S, Klepac P, Jit M, Diamond C, Gibbs H, van Zandvoort

- 822 K, Funk S, Eggo RM (2020) Feasibility of controlling covid-19 outbreaks by  
823 isolation of cases and contacts. *The Lancet Global Health* 8(4):e488–e496  
824 [10](#)
- 825 Henriques R, Lobo V, Bação F (2012) Spatial clustering using hierarchical  
826 som. *Applications of Self-Organizing Maps* pp 231–250 [13](#)
- 827 Hernández-Orallo E, Manzoni P, Calafate CT, Cano JC (2020) Evaluating how  
828 smartphone contact tracing technology can reduce the spread of infectious  
829 diseases: the case of covid-19. *IEEE Access* [2](#), [5](#), [6](#), [27](#)
- 830 Keeling MJ, Rohani P (2011) *Modeling infectious diseases in humans and*  
831 *animals*. Princeton University Press [6](#), [10](#)
- 832 Lang JC, De Sterck H, Kaiser JL, Miller JC (2018) Analytic models for sir  
833 disease spread on random spatial networks. *Journal of Complex Networks*  
834 6(6):948–970 [4](#)
- 835 Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J (2020) Substantial  
836 undocumented infection facilitates the rapid dissemination of novel coron-  
837 avirus (sars-cov2). *Science* [10](#)
- 838 Mahsin M, Deardon R, Brown P (2020) Geographically dependent individual-  
839 level models for infectious diseases transmission. *Biostatistics* [4](#)
- 840 Martínez-Beneito MA, Mateu J, Botella-Rocamora P (2020) Spatio-temporal  
841 small area surveillance of the covid-19 pandemics. *arXiv preprint*  
842 *arXiv:201103938* [4](#)
- 843 Martínez-Martin N, Wieten S, Magnus D, Cho MK (2020) Digital contact  
844 tracing, privacy, and public health. *Hastings Center Report* 50(3):43–46 [3](#)
- 845 Müller M, Derlet PM, Mudry C, Aeppli G (2020) Testing of asymptomatic  
846 individuals for fast feedback-control of covid-19 pandemic. *Physical biology*  
847 17(6):065,007 [3](#)
- 848 Park J, Chang W, Choi B (2021) An interaction neyman-scott point process  
849 model for coronavirus disease-19. *arXiv preprint arXiv:210202999* [5](#)
- 850 Reichert L, Brack S, Scheuermann B (2020) Privacy-preserving contact tracing  
851 of covid-19 patients. *IACR Cryptol ePrint Arch* 2020:375 [3](#)
- 852 Rezaei M, Azarmi M (2020) Deepsocial: Social distancing monitoring and in-  
853 fection risk assessment in covid-19 pandemic. *Applied Sciences* 10(21):7514  
854 [12](#), [18](#)
- 855 Salathé M (2018) Digital epidemiology: what is it, and where is it going? *Life*  
856 *sciences, society and policy* 14(1):1 [2](#)
- 857 Simmerman JM, Suntarattiwong P, Levy J, Gibbons RV, Cruz C, Shaman  
858 J, Jarman RG, Chotpitayasunondh T (2010) Influenza a Virus Contam-  
859 ination of Common Household Surfaces during the 2009 Influenza A  
860 (H1N1) Pandemic in Bangkok, Thailand: Implications for Contact Trans-  
861 mission. *Clinical Infectious Diseases* 51(9):1053–1061, DOI 10.1086/656581,  
862 URL <https://doi.org/10.1086/656581>, [https://academic.oup.com/  
cid/article-pdf/51/9/1053/827213/51-9-1053.pdf](https://academic.oup.com/cid/article-pdf/51/9/1053/827213/51-9-1053.pdf) [3](#)
- 864 Souza RC, Neill DB, Assunção RM, Meira W (2019) Identifying high-risk areas  
865 for dengue infection using mobility patterns on twitter. *Online Journal of*  
866 *Public Health Informatics* 11(1) [4](#)

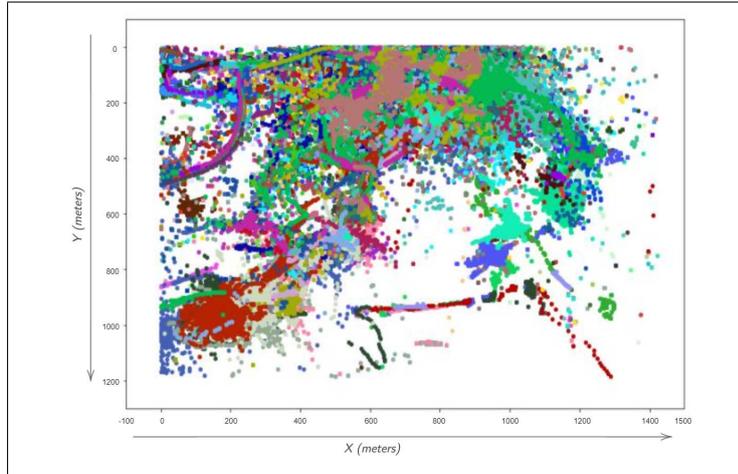
- 867 Takács B, Hadjimichael Y (2019) High order discretization methods for spatial  
868 dependent sir models. arXiv preprint arXiv:190901330 [4](#)
- 869 Tsai TC, Chan HH (2015) Nccu trace: Social-network-aware mobility trace.  
870 IEEE Communications Magazine 53(10):144–149 [5](#), [15](#)
- 871 Van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A,  
872 Williamson BN, Tamin A, Harcourt JL, Thornburg NJ, Gerber SI, et al  
873 (2020) Aerosol and surface stability of sars-cov-2 as compared with sars-  
874 cov-1. New England Journal of Medicine 382(16):1564–1567 [3](#)
- 875 Vesanto J, Alhoniemi E (2000) Clustering of the self-organizing map. IEEE  
876 Transactions on neural networks 11(3):586–600 [13](#), [14](#)
- 877 Wang P, Fu Y, Liu G, Hu W, Aggarwal C (2017) Human mobility synchro-  
878 nization and trip purpose detection with mixture of hawkes processes. In:  
879 Proceedings of the 23rd ACM SIGKDD international conference on knowl-  
880 edge discovery and data mining, pp 495–503 [2](#)
- 881 WHO (2021) WHO Coronavirus Disease (COVID-19) Dashboard — WHO  
882 Coronavirus Disease (COVID-19) Dashboard. URL [https://covid19.who.](https://covid19.who.int/)  
883 [int/](https://covid19.who.int/) [2](#)
- 884 Zheng Y (2015) Trajectory data mining: an overview. ACM Transactions on  
885 Intelligent Systems and Technology (TIST) 6(3):1–41 [3](#)
- 886 Zheng Y, Fu H, Xie X, Ma W, Li Q (2011) Geolife gps trajectory dataset-user  
887 guide. microsoft research [26](#)
- 888 Zhou C, Yuan W, Wang J, Xu H, Jiang Y, Wang X, Wen QH, Zhang P (2020)  
889 Detecting suspected epidemic cases using trajectory big data. arXiv preprint  
890 arXiv:200400908 [2](#)

## 891 5 Appendix: NCCU data

892 This section provides details related to the study area and the selected dataset.  
893 Figure 18 depicts coordinates of the study area, whereas Figure 19 illustrates  
894 the complete dataset of all 115 individuals for the period of 15 days where  
895 each user is shown with a different colour.

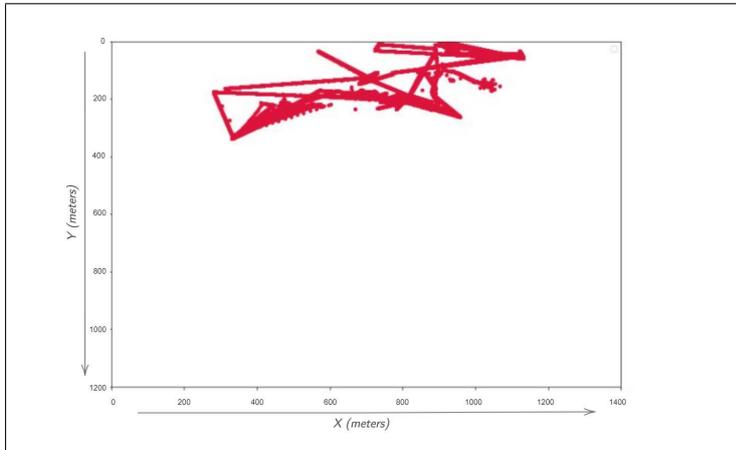


**Fig. 18** Coordinates of the study area

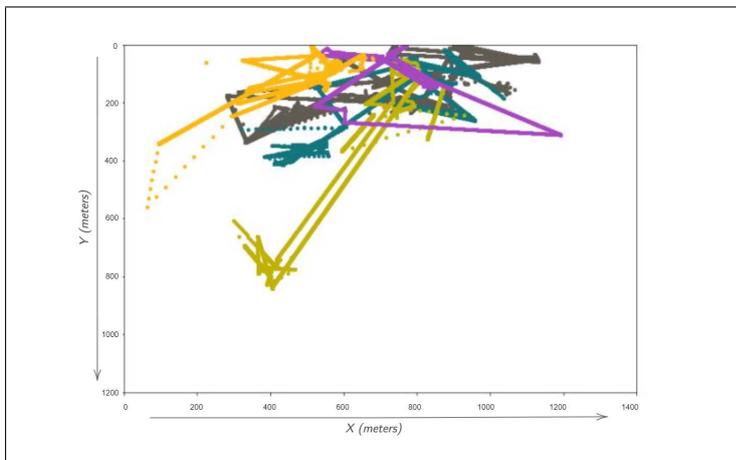


**Fig. 19** Extent of recorded dataset

896 Figure 20 illustrates mobility trajectories of a single user for a period of  
897 one day, and Figure 21 shows 1-day movement for 5 users.

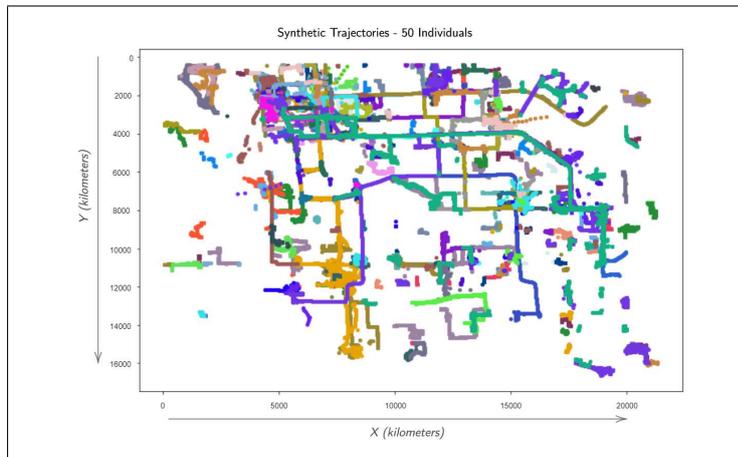


**Fig. 20** Mobility trajectory of a single user for a 1-day period



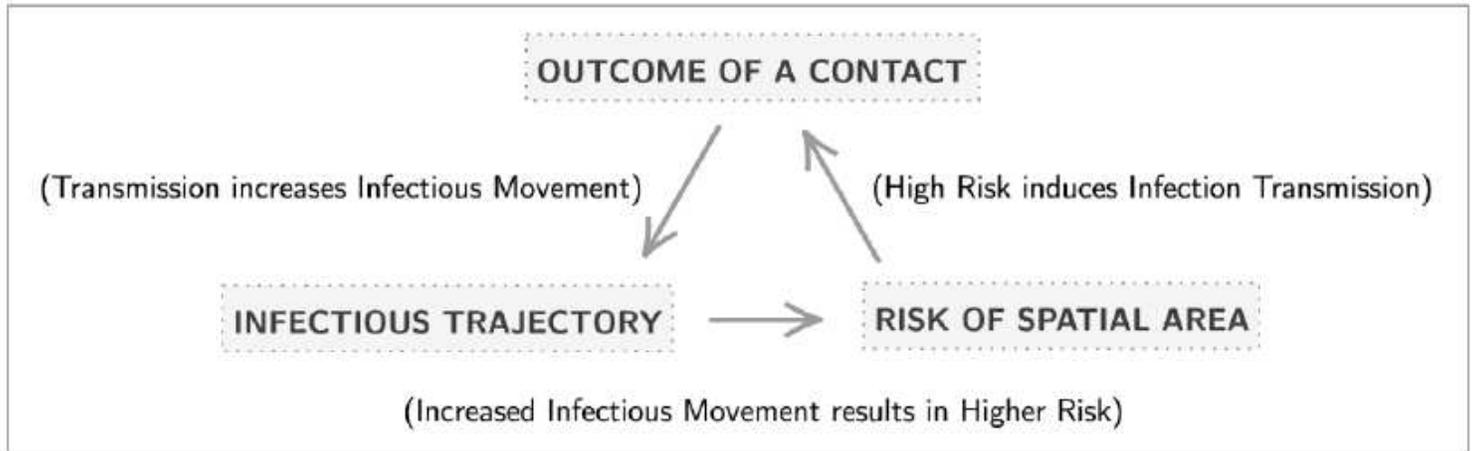
**Fig. 21** Mobility trajectory of five users for a 1-day period

898 Figure 22 shows self-generated mobility trajectories of all 50 users for a  
899 period of 15 days.



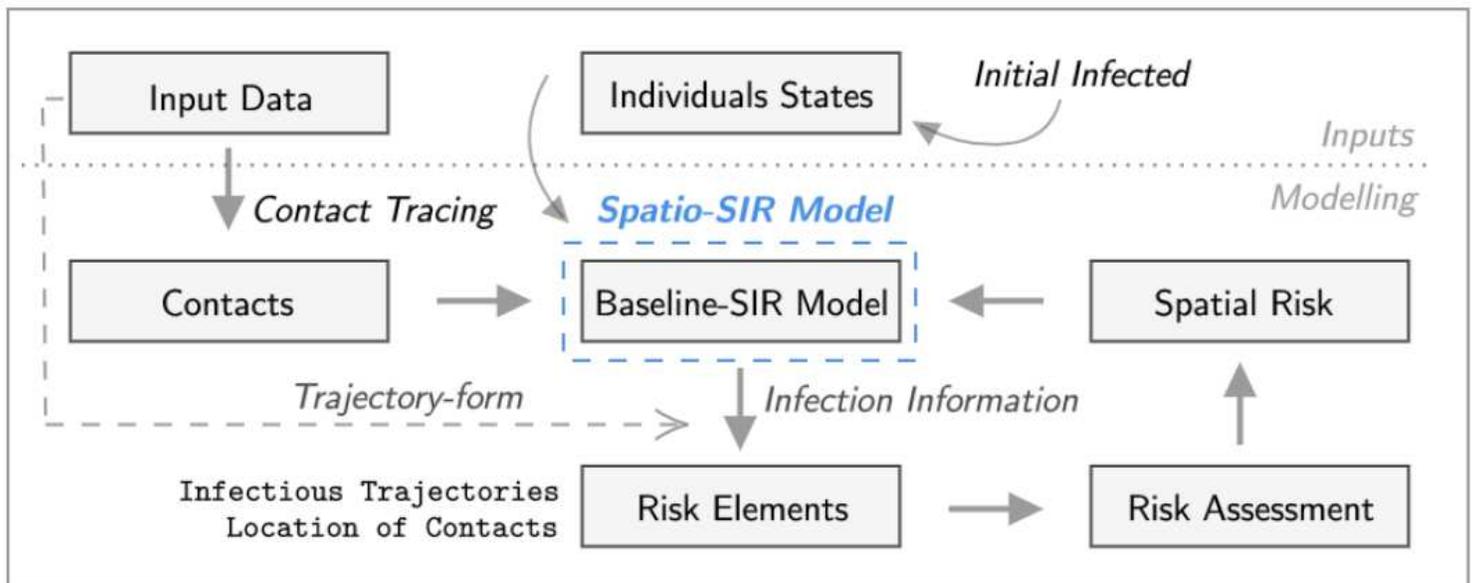
**Fig. 22** Geolife data based self simulated mobility trajectories of 50 users for complete data period of 15 days

# Figures



**Figure 1**

Relation of temporally varying spatial risk affecting epidemic model and vice versa



**Figure 2**

Summarized workflow of overall methodology

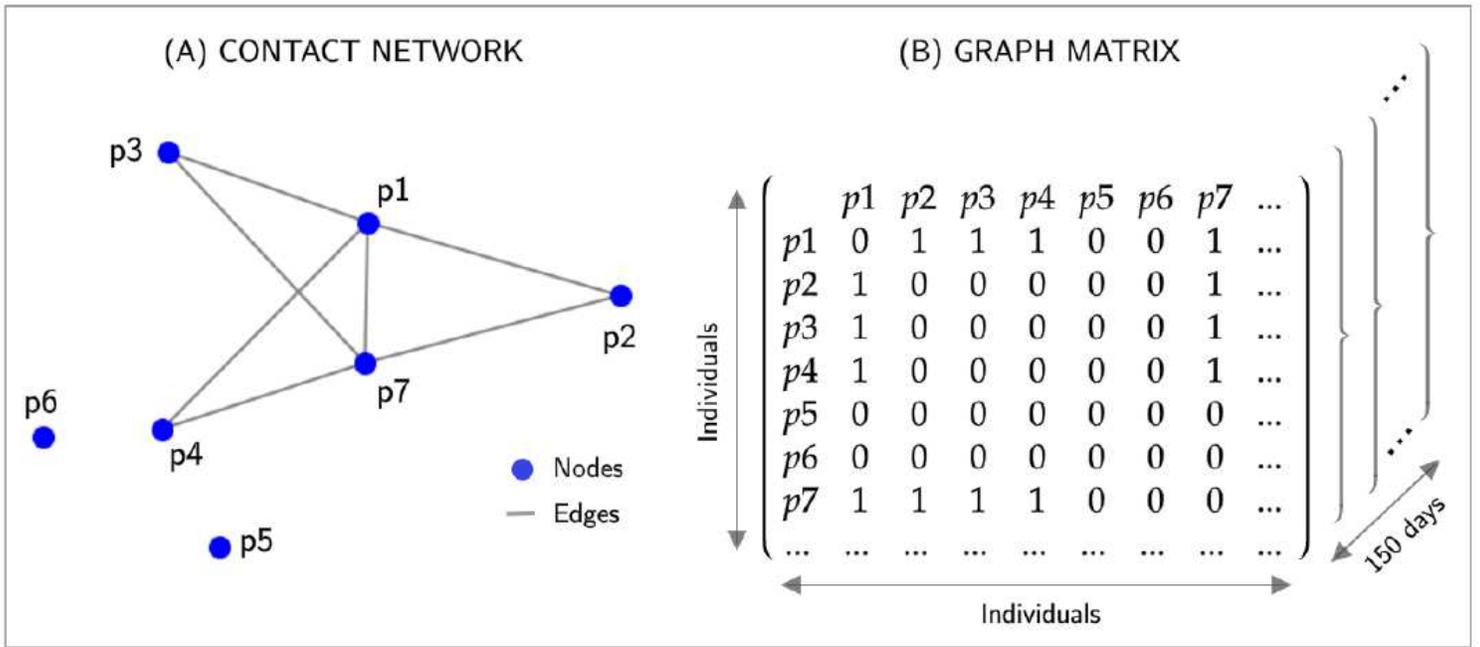


Figure 3

A toy example: contacts of a single day in the form of contact network and graph matrix

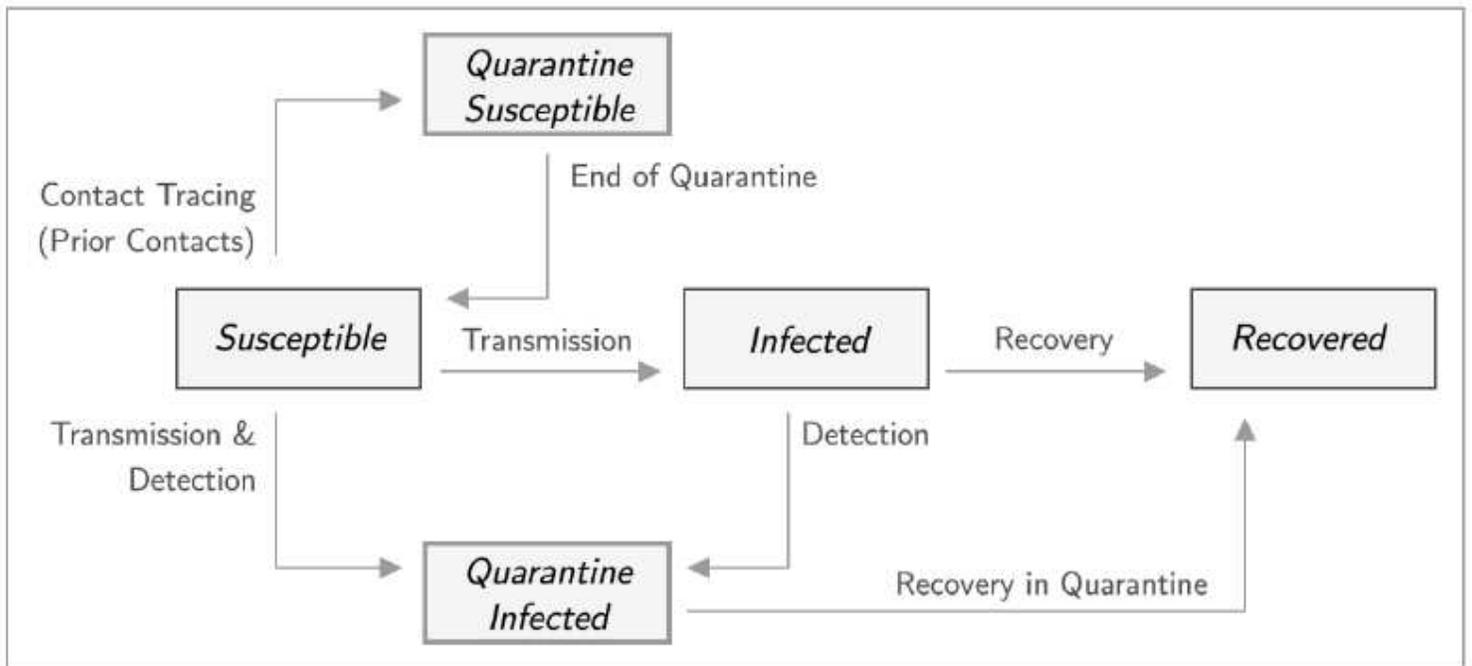
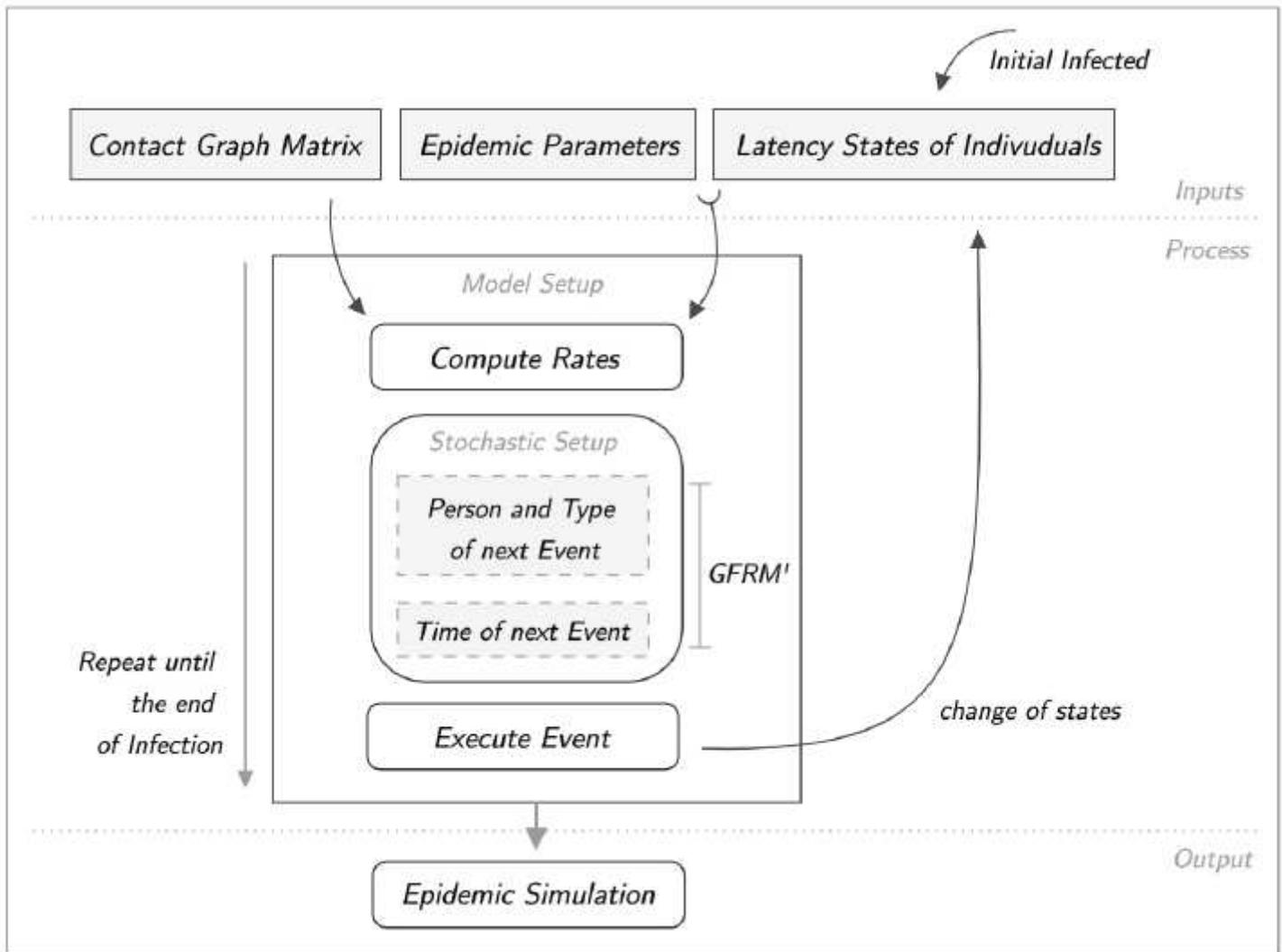


Figure 4

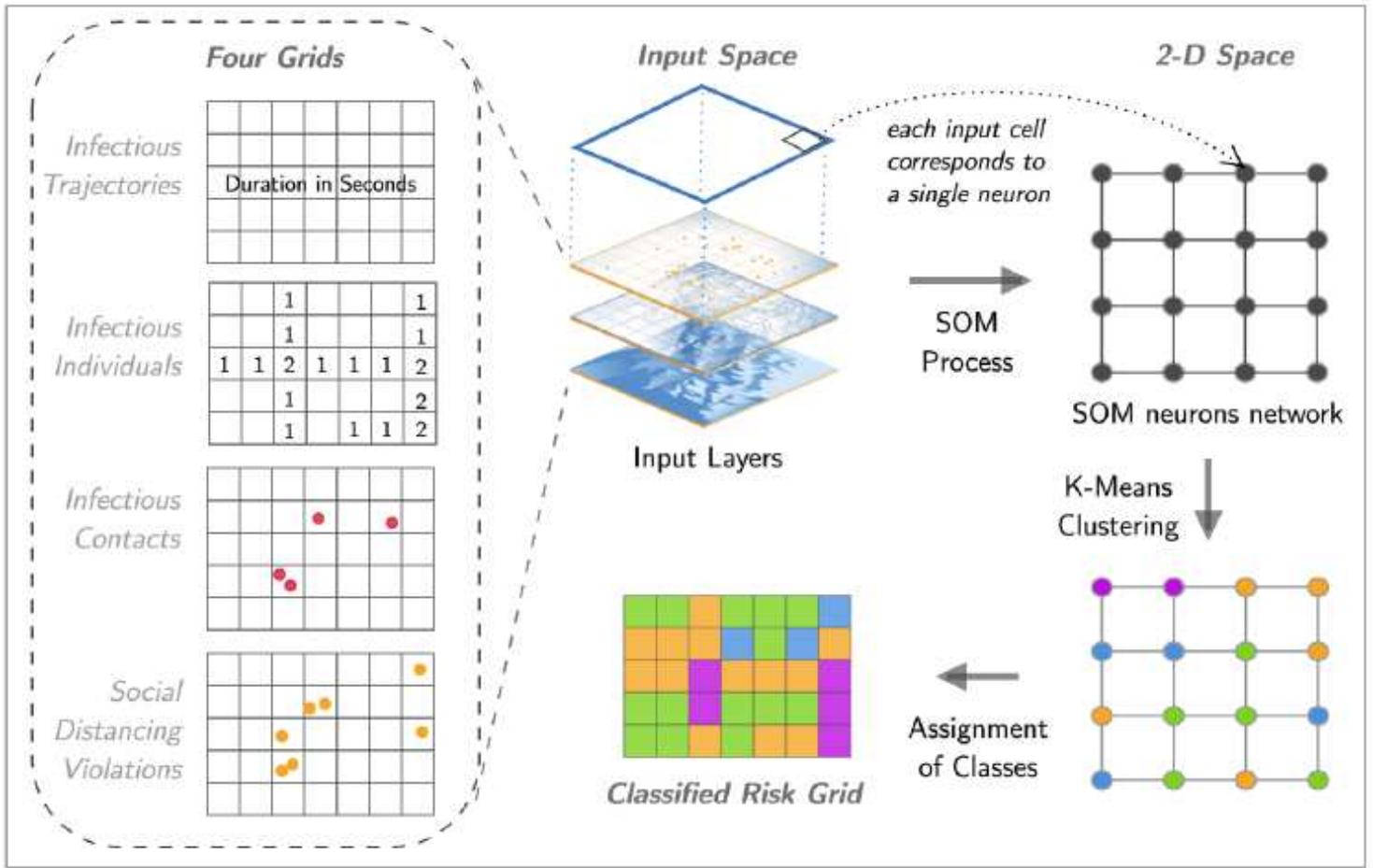
Overview of compartments and possible transfers between them



**Figure 5**

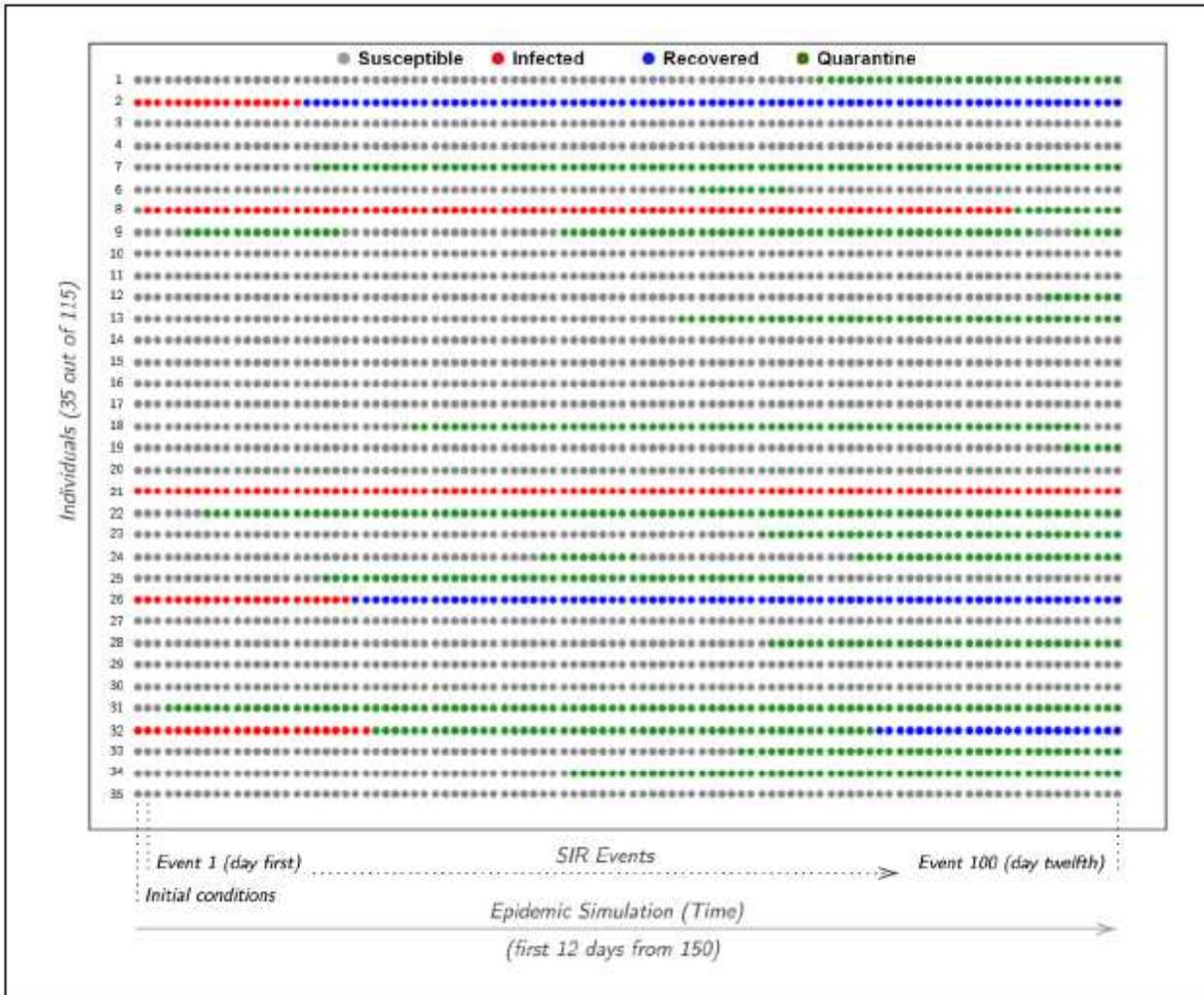
Process flow of an event-driven stochastic SIR model





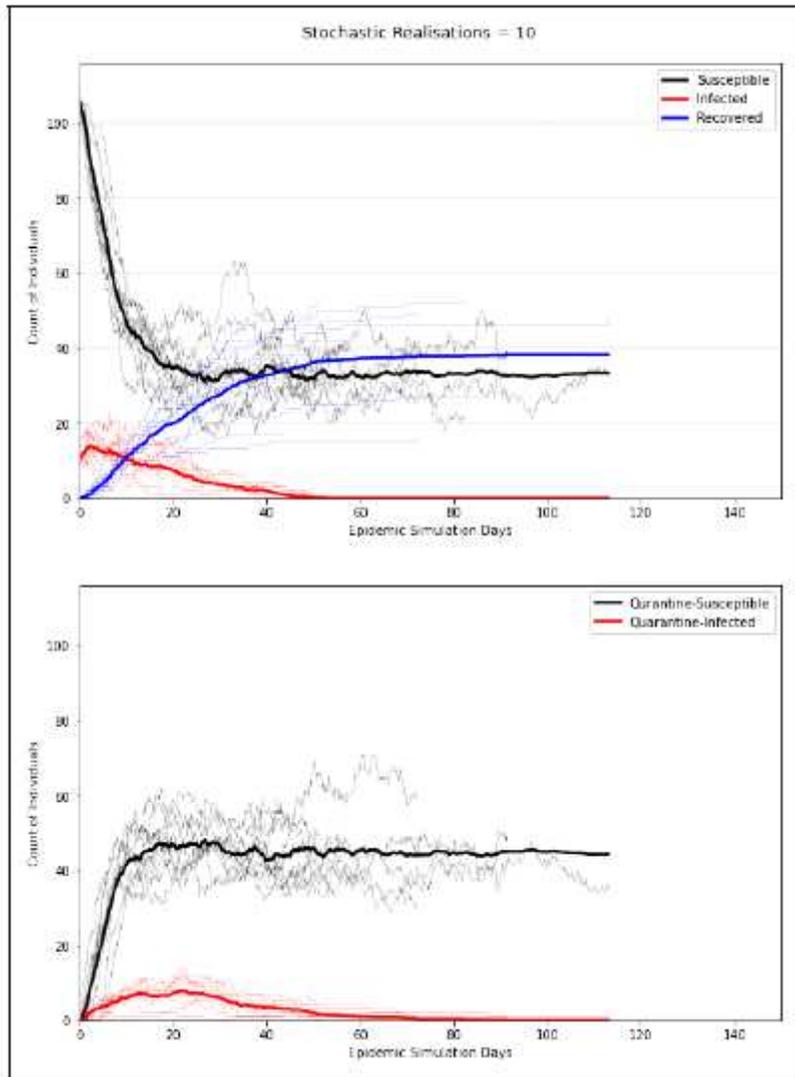
**Figure 7**

Unsupervised classification workflow using SOM and K-means



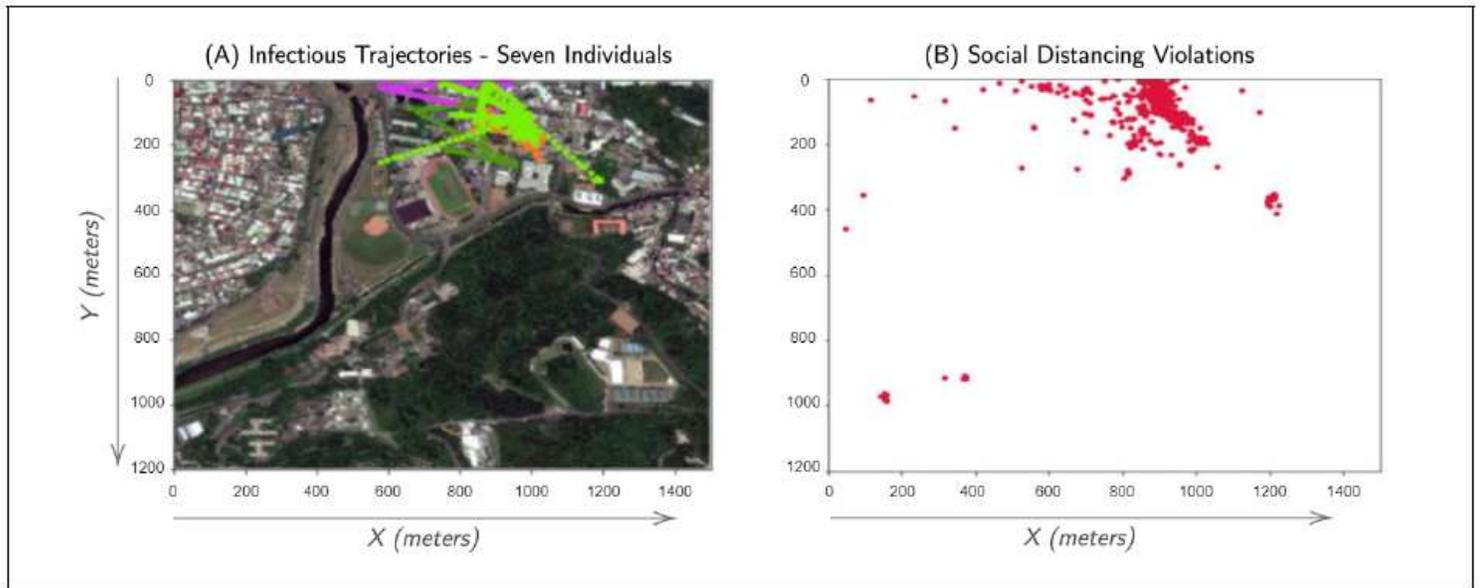
**Figure 8**

Individual-level change in latency of 35 out of 115 total individuals is shown based on the SIR events as model runs forward. Each row belongs to a single individual, where the compartment they belong to at an instance of time is represented column-wise chrono-logically from left to right. There is only one event per column with multiple events per day, where figure illustrates first 100 events from the initial 12 days of epidemic.



**Figure 9**

Output of a disease outbreak scenario by baseline-SIR model. (Top) presents trends related to counts of Susceptible, Infected and Recovered, whereas (Bottom) illustrates count of individuals in Quarantine related compartments. Each of the 10 realizations of stochastic model is shown (light in color), with their average curves represented with (dark bold) lines. Count of total population is 115 which are represented over the Y-axis.



**Figure 10**

Risk basis of infectious trajectories and location of contacts from one complete day during an epidemic. In the sample shown in Sub-Figure (A), there are total of 7 infected individuals from that day. Sub-Figure (B) shows locations of all contacts from the same day. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

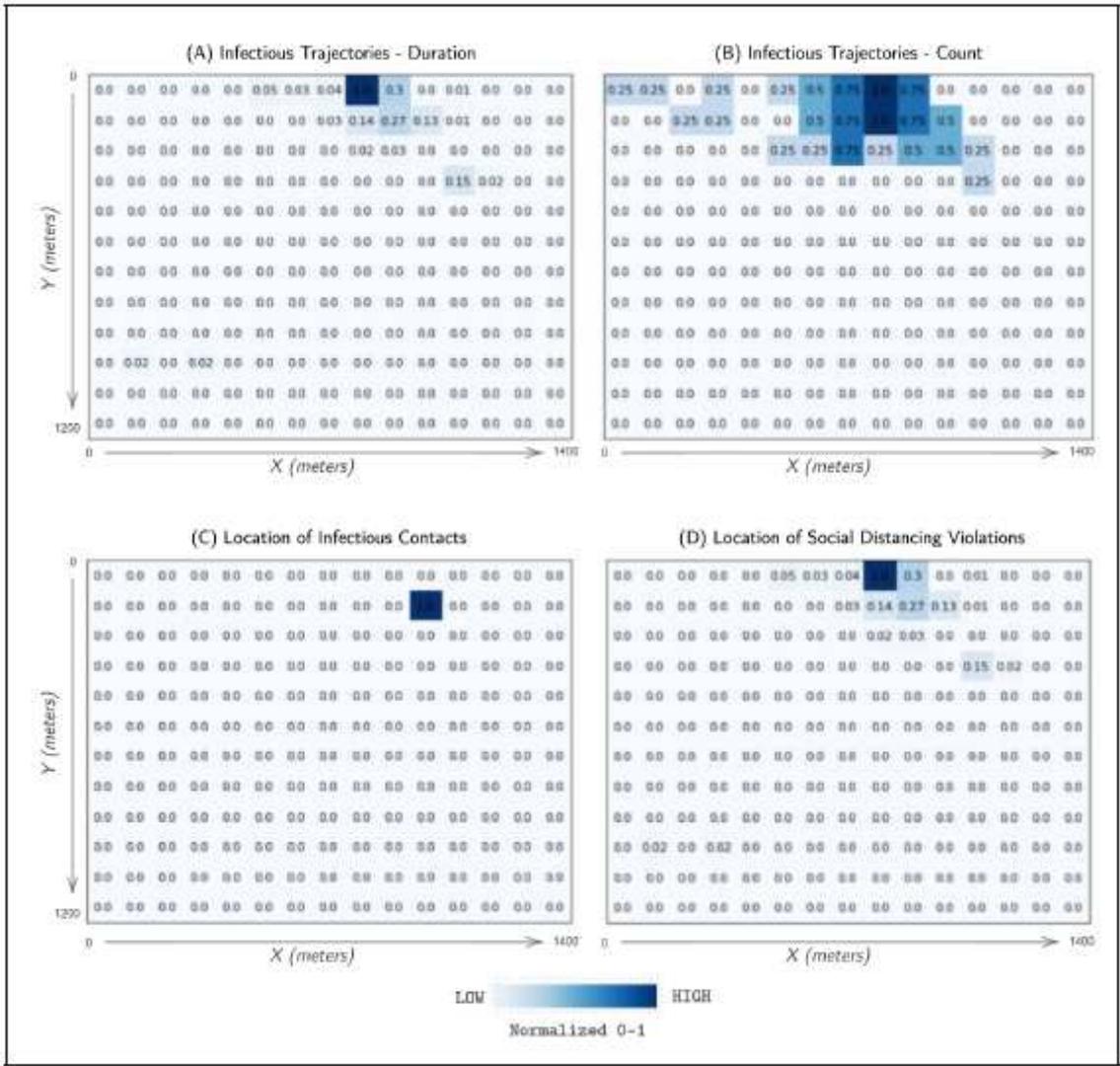
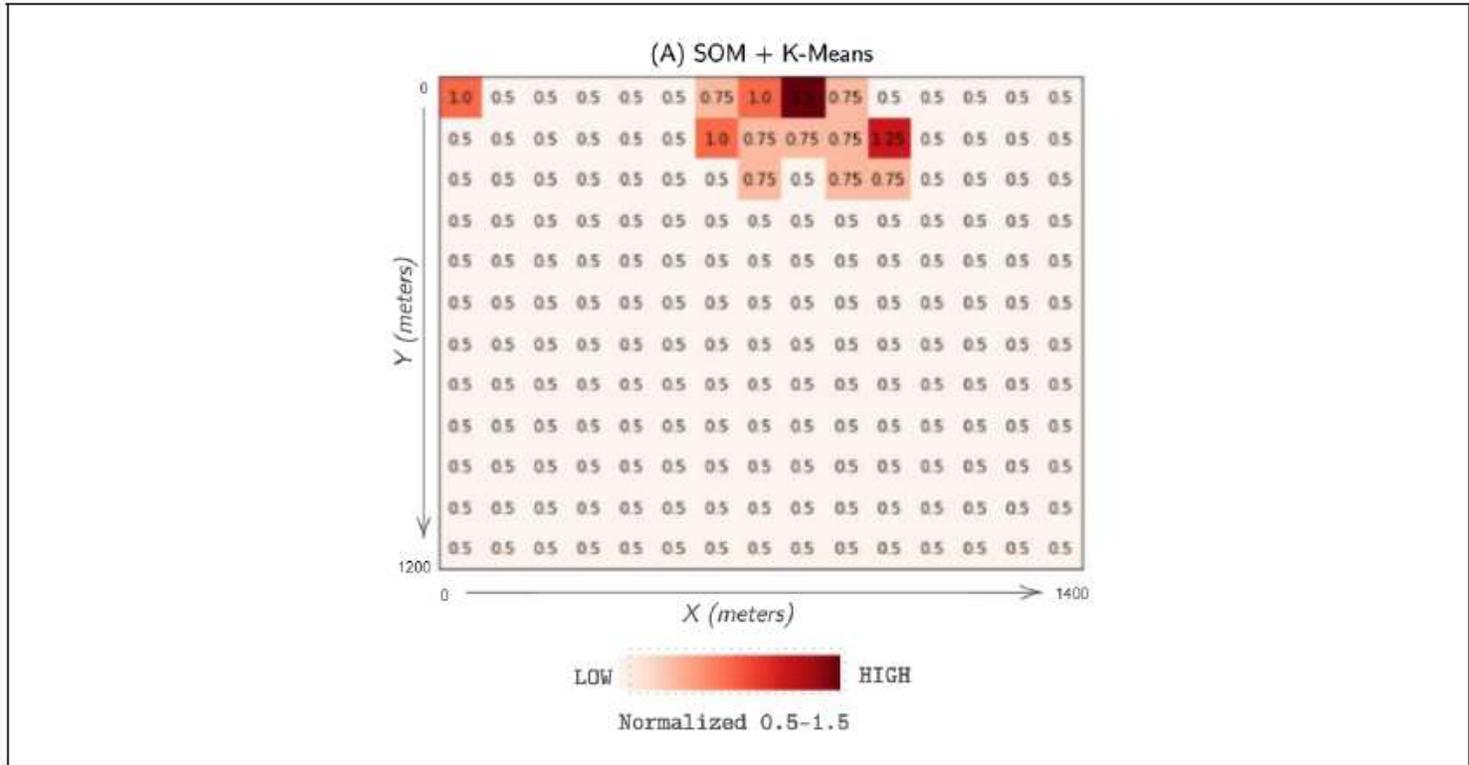


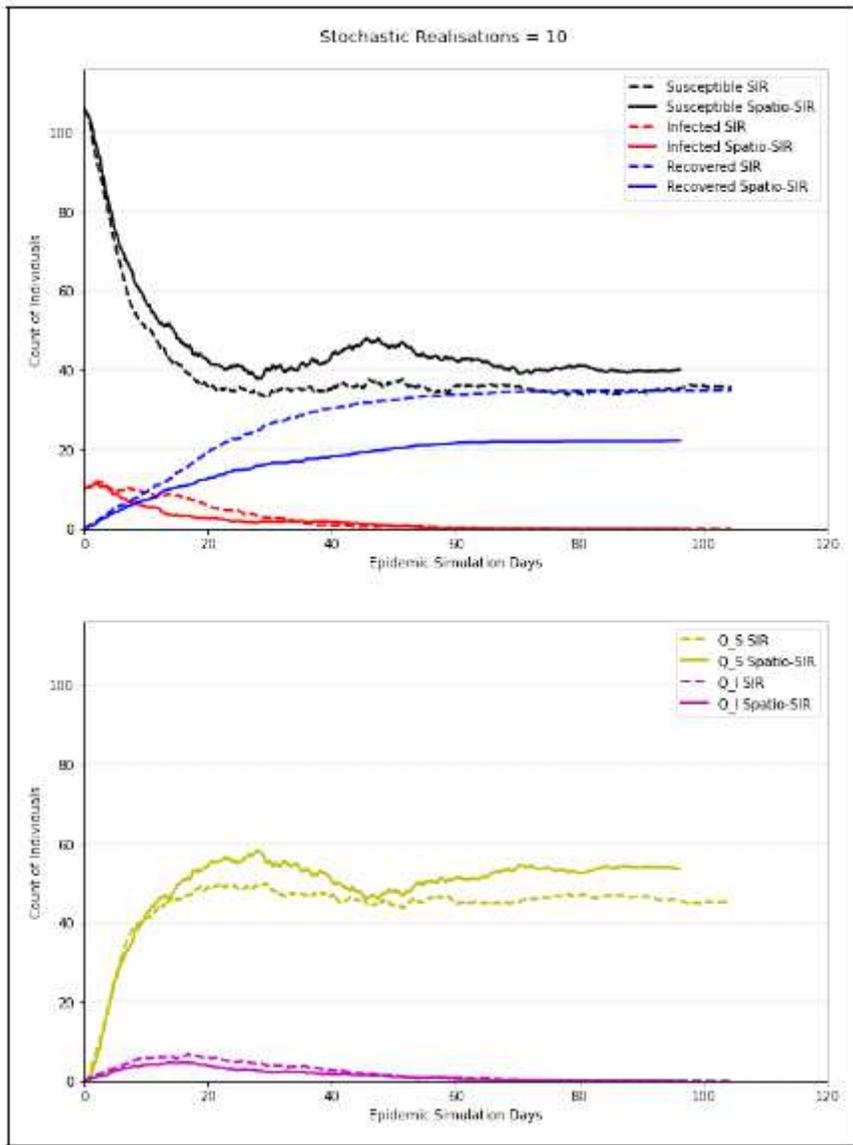
Figure 11

Grid-based risk computed out of risk basis. Sub-Figure (A) and (B) refer to grid-based representation of duration and count of trajectories shown in Figure 10-(A). Sub-Figure (C) depicts location of an infectious contact occurred on the same day, whereas Sub-Figure (D) presents translation of Figure 10-(B) into grid form. All values are normalized to the range of 0 to 1, where 1 refers to highest risk.



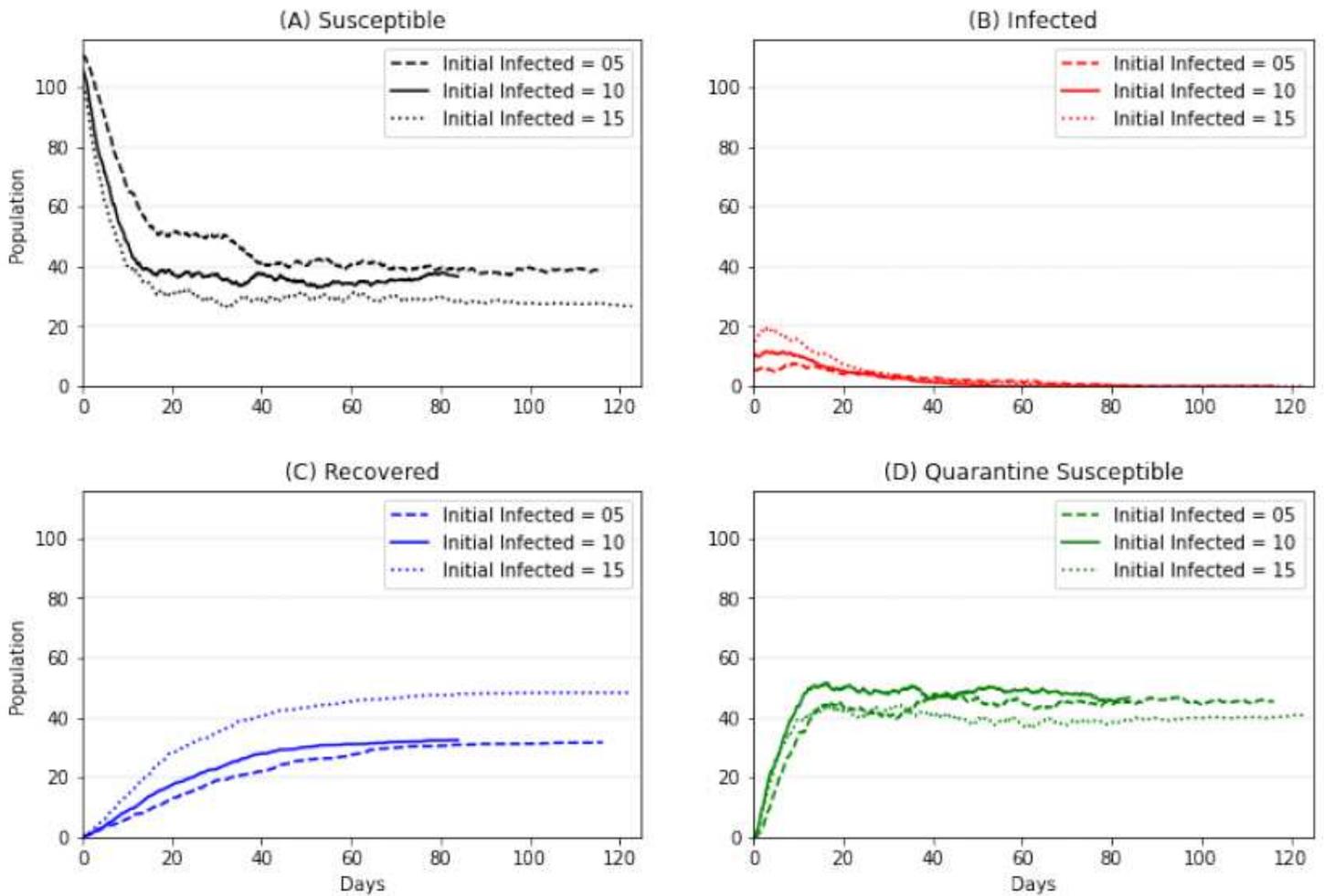
**Figure 12**

Combining risk from multiple grids shown in Figure 11 into a single grid output using SOM and K-means. This integration is executed in unsupervised manner through the implementation of SOM followed by K-means. Risk scores are computed in the range of [0.5,1.5], where 1 refers to the previous normal (existing SIR model with a constant spatial risk and all contacts being of equal nature)



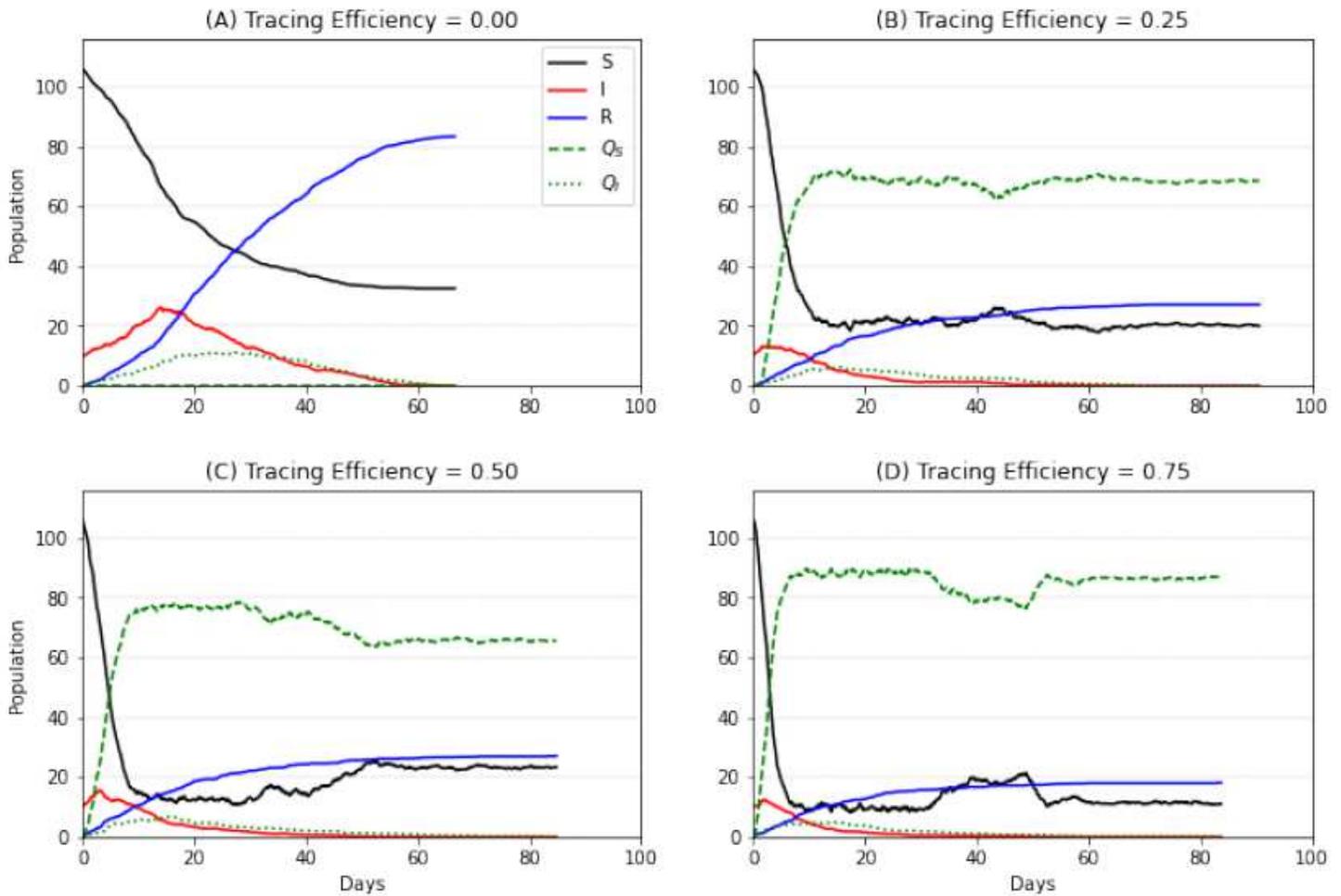
**Figure 13**

Comparing average of 10 stochastic realizations of a disease outbreak scenario from baseline-SIR (dashed) and spatio-SIR (solid). (Top) presents trends related to count of Susceptible, Infected and Recovered, whereas (Bottom) illustrates counts of individuals in Quarantine related compartments. Count of total population is 115 which are represented over Y-axis.



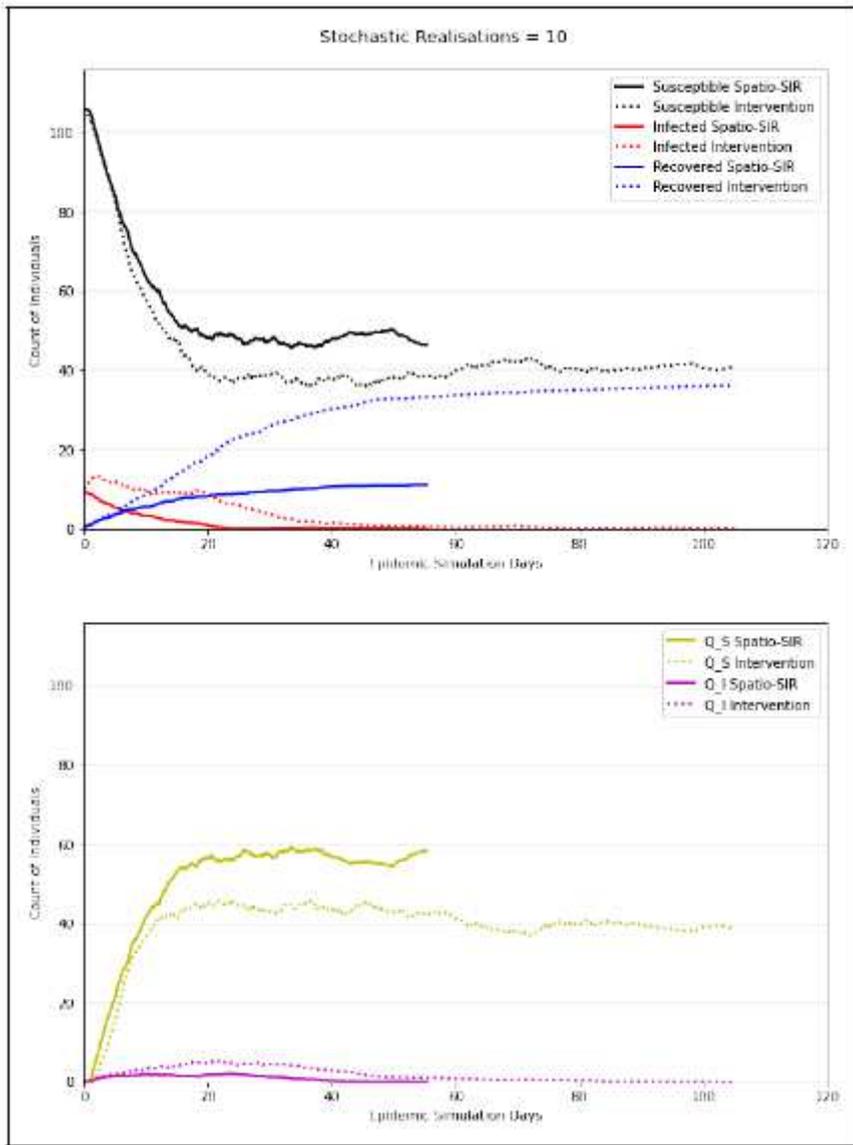
**Figure 14**

Comparing the average of 10 stochastic realizations (of baseline-SIR model) with varying count of Initial Infected  $I_0$ , to observe their effect on the overall disease outbreak. Three scenarios of  $I_0 = 5/10/15$  are shown with a population size of 115. Subplots (A,B,C and D) shows comparison of Susceptible, Infected, Recovered and Quarantine Susceptible, respectively. Due to stochasticity, final duration of epidemic varies depending on the overall spread of infection.



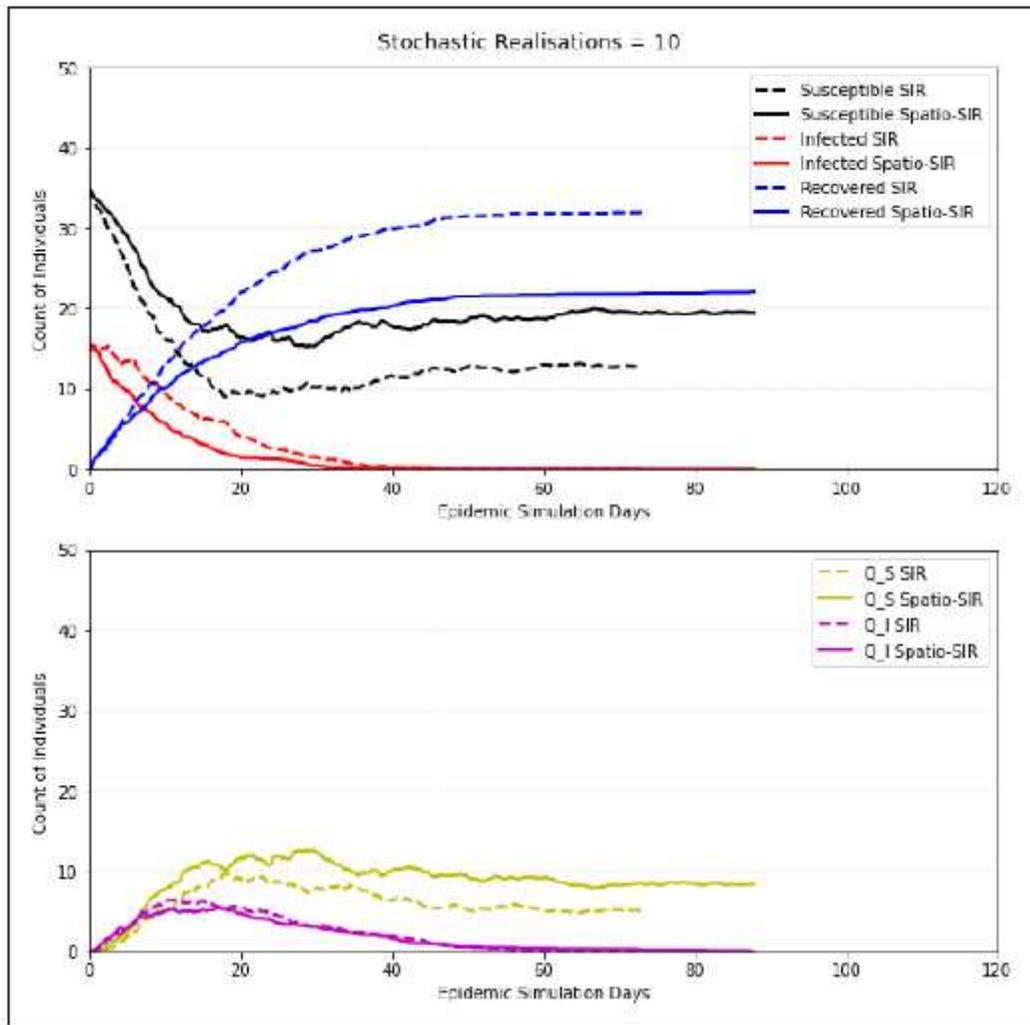
**Figure 15**

Comparing the average of 10 stochastic realizations (of baseline-SIR model) with varying Tracing Efficiency  $q'$  to observe their effect on the overall disease outbreak (population size = 115). Subplots (A,B,C and D) present four cases of  $q'=0, 0.25, 0.50$  and  $0.75$ , respectively, where 1 means 100% backward tracing.



**Figure 16**

Comparing average of 10 stochastic realizations of a disease outbreak scenario from spatio-SIR (solid) and a case of Intervention - spatio-SIR model with spatial risk of 1.5 from day 11 to 20 (dotted). (Top) presents trends related to counts of Susceptible, Infected and Recovered, whereas (Bottom) illustrates counts of individuals in Quarantine related compartments. Count of total population is 115 which are represented over Y-axis.



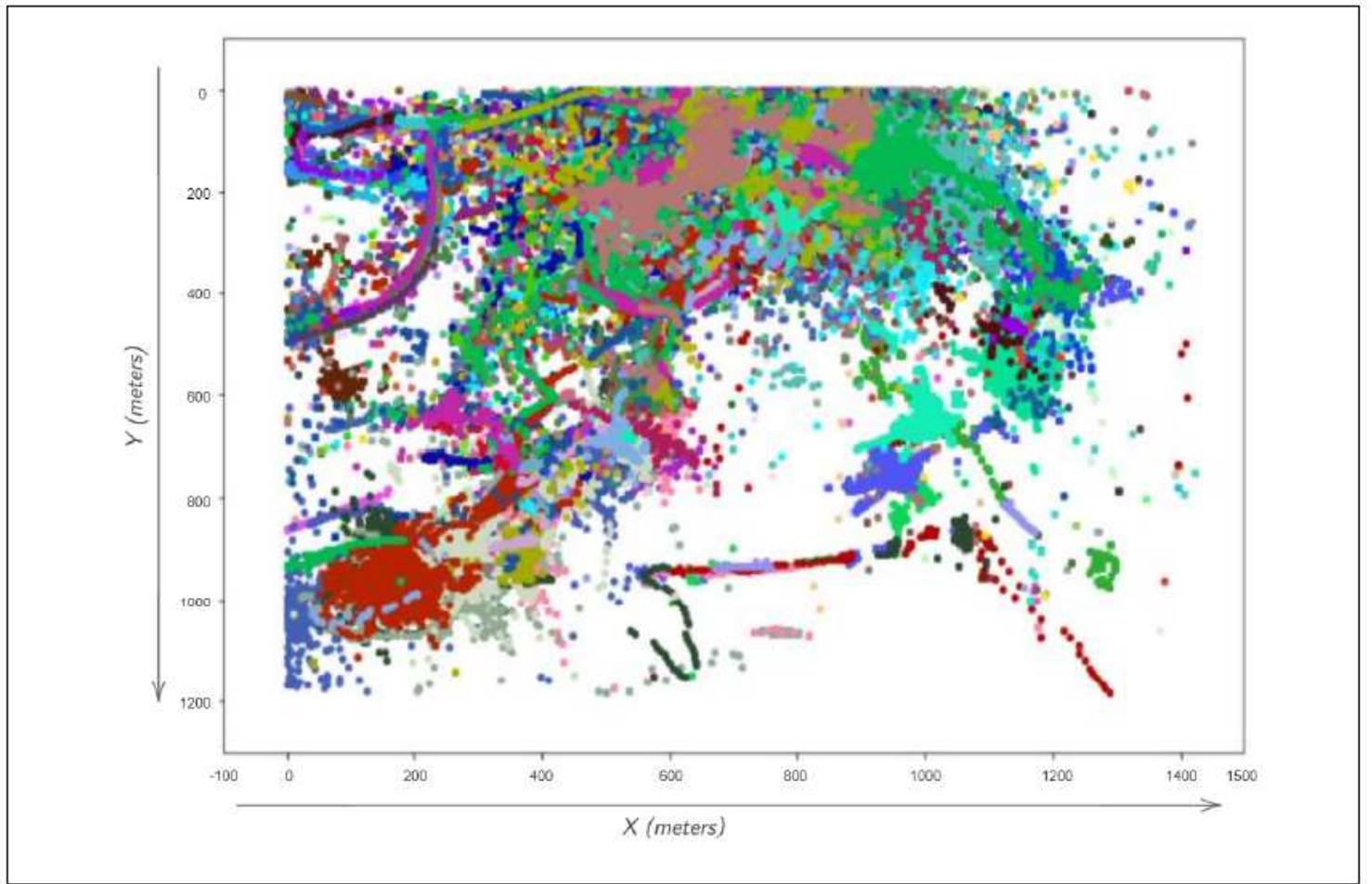
**Figure 17**

Comparing average of 10 stochastic realizations of a disease outbreak scenario from baseline-SIR (dashed) and spatio-SIR (solid). (Top) presents trends related to counts of Susceptible, Infected and Recovered, whereas (Bottom) illustrates counts of individuals in Quarantine related compartments. Count of total population is 50 which are represented over Y-axis.



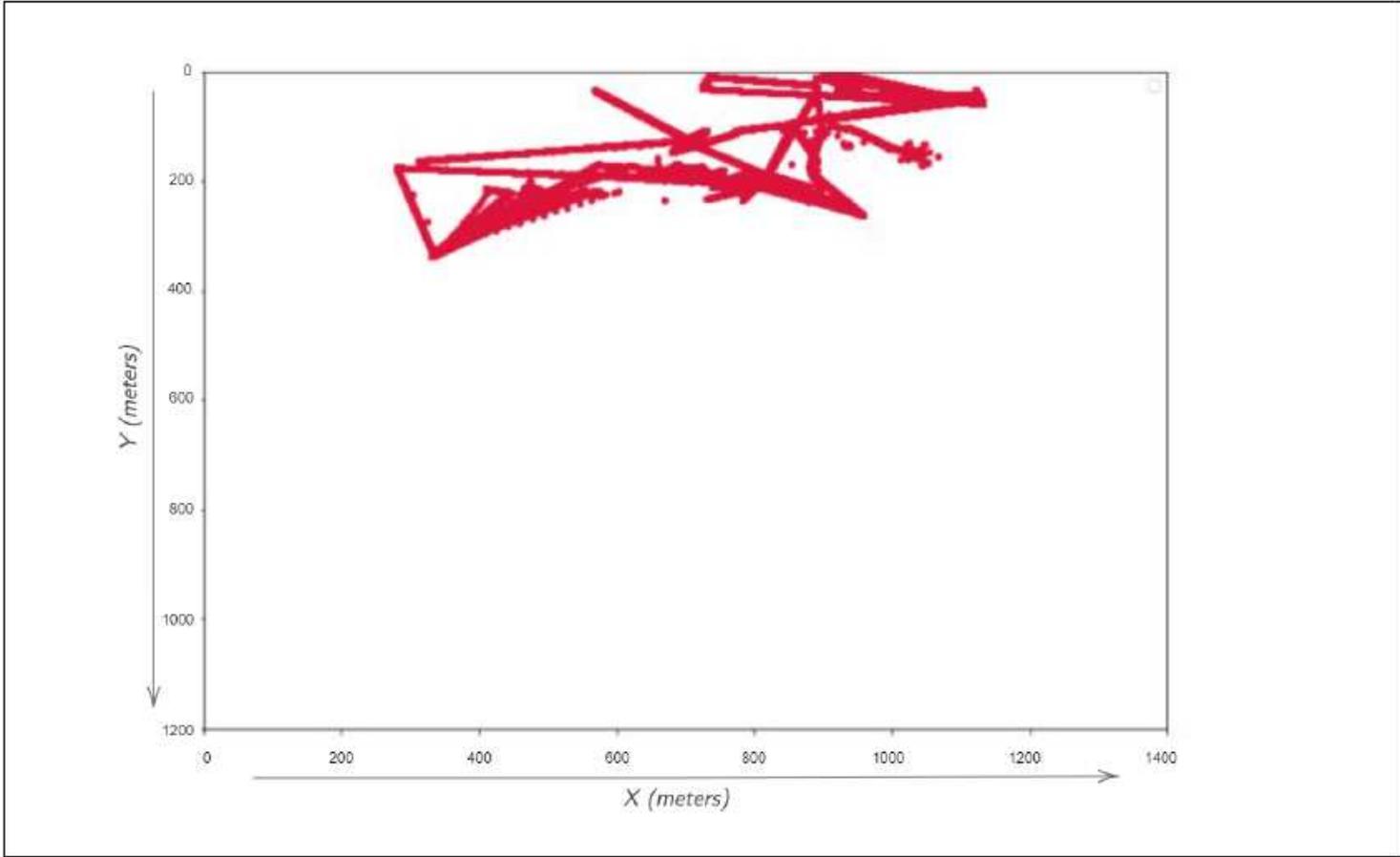
**Figure 18**

Coordinates of the study area Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



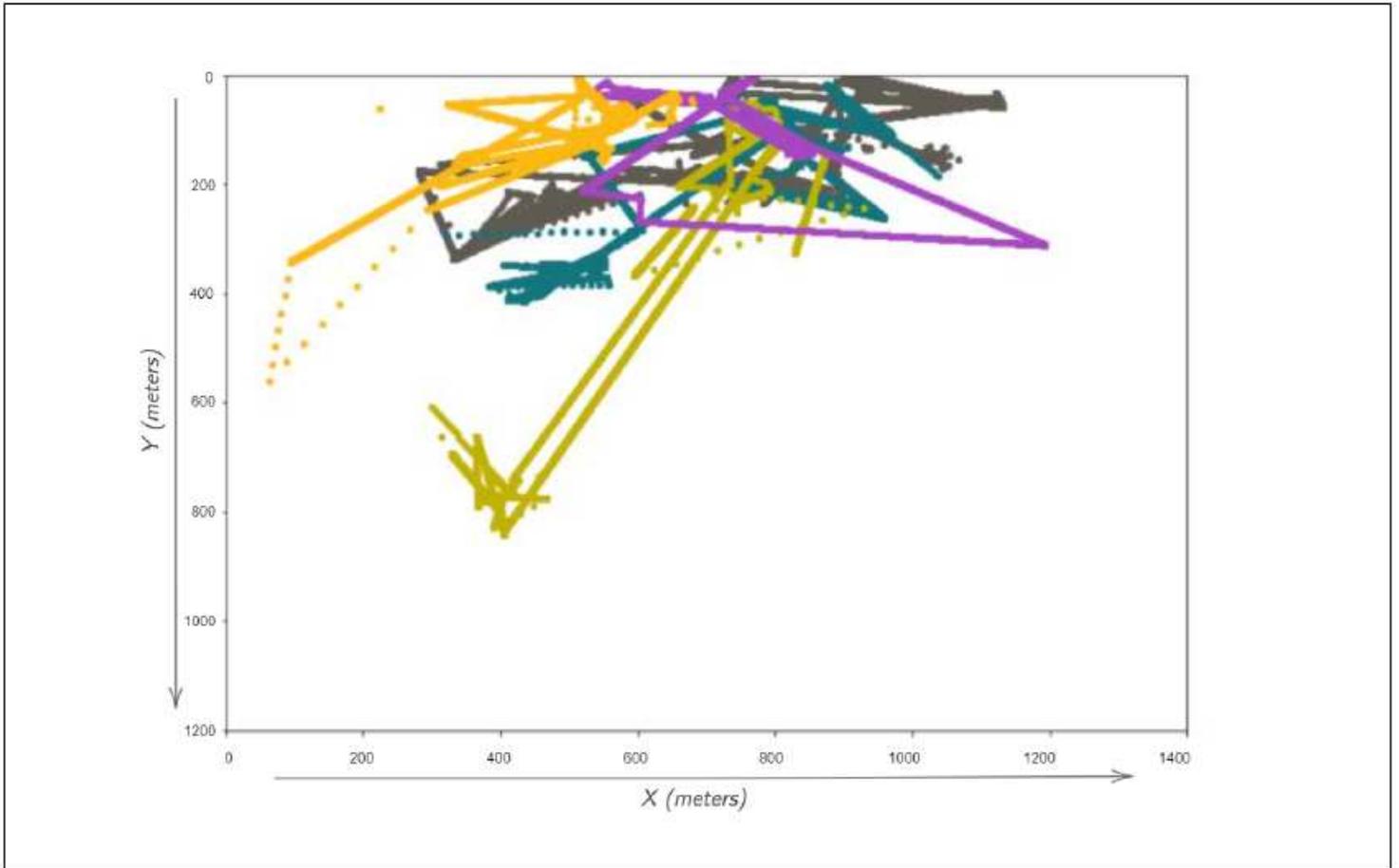
**Figure 19**

Extent of recorded dataset



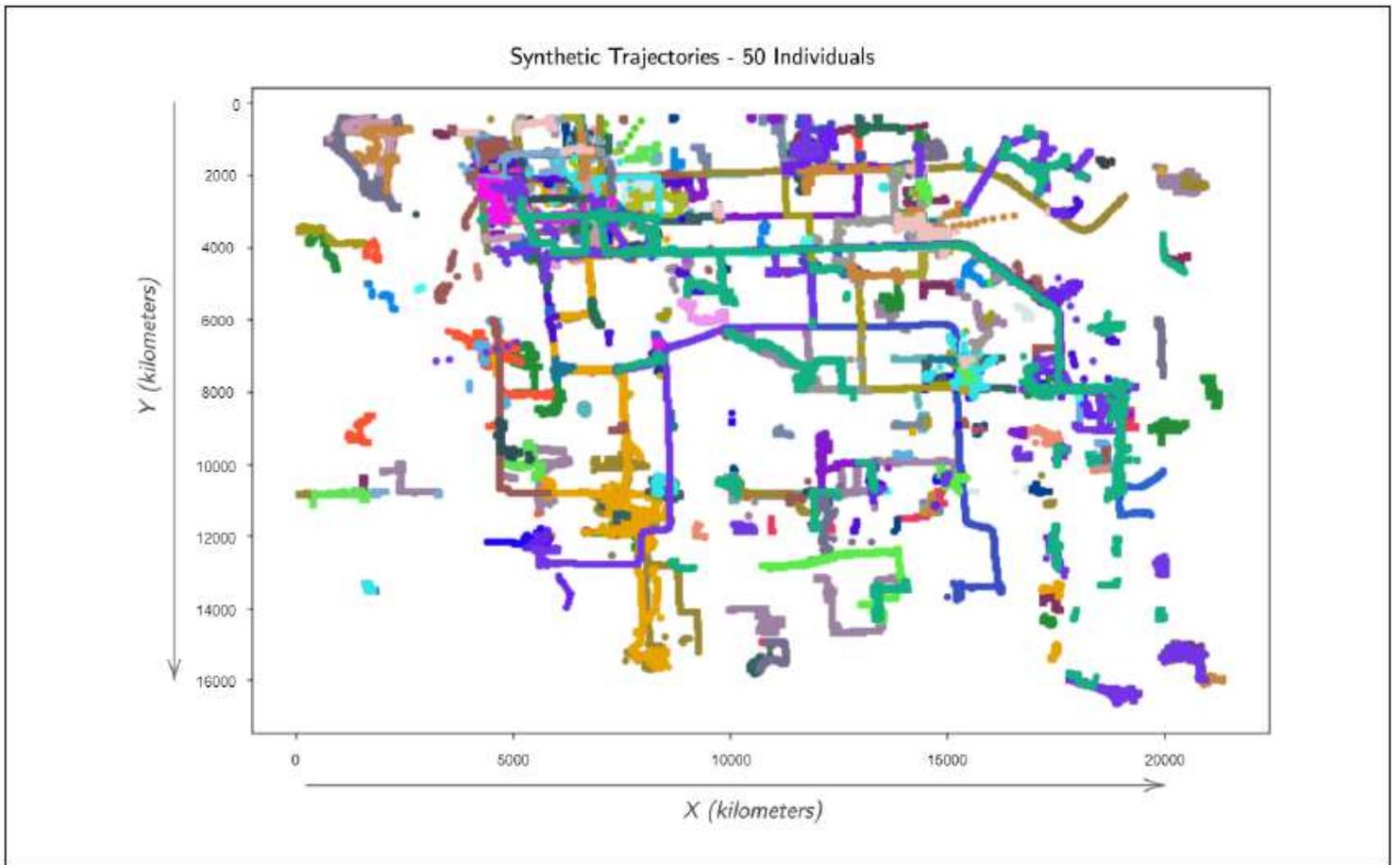
**Figure 20**

Mobility trajectory of a single user for a 1-day period



**Figure 21**

Mobility trajectory of five users for a 1-day period



**Figure 22**

Geolife data based self simulated mobility trajectories of 50 users for complete data period of 15 days