Scientific Research Publishing

Open Journal of Modelling and Simulation, 2023, 11, 99-116
https://www.scirp.org/journal/ojmsi
ISSN Online: 2327-4026
ISSN Print: 2327-4018

"stppSim": A Novel Analytical Tool for Creating Synthetic Spatio-Temporal Point Data

Monsuru Adepeju

Crime and Well-Being Big Data Centre, Manchester Metropolitan University, Manchester, UK
Email: m.adepeju@mmu.ac.uk

How to cite this paper: Adepeju, M. (2023) "stppSim": A Novel Analytical Tool for Creating Synthetic Spatio-Temporal Point Data. *Open Journal of Modelling and Simulation*, **11**, 99-116.
https://doi.org/10.4236/ojmsi.2023.114007

Received: August 23, 2023
Accepted: October 9, 2023
Published: October 12, 2023

Abstract

In crime science, understanding the dynamics and interactions between crime events is crucial for comprehending the underlying factors that drive their occurrences. Nonetheless, gaining access to detailed spatiotemporal crime records from law enforcement faces significant challenges due to confidentiality concerns. In response to these challenges, this paper introduces an innovative analytical tool named "stppSim," designed to synthesize fine-grained spatiotemporal point records while safeguarding the privacy of individual locations. By utilizing the open-source R platform, this tool ensures easy accessibility for researchers, facilitating download, re-use, and potential advancements in various research domains beyond crime science.

Keywords

Open-Source, Synthetic Data, Crime, Spatio-Temporal Patterns, Data Privacy

1. Introduction

Understanding crime dynamics holds immense significance in shaping effective policies and fostering safer communities [1]. By delving into the complex interplay of factors that drive criminal behavior and patterns, we gain insights that can guide targeted interventions and law enforcement strategies. This importance lies in its potential to prevent crime. The analysis of temporal patterns and spatial concentrations, along with the intricate interconnection of these dimensions in criminal behaviour, empowers law enforcement entities to allocate resources strategically and deploy preventive measures proactively [2] [3]. This, in turn, reduces the opportunity for criminal acts to occur.

Nonetheless, the ongoing advancements in digital data acquisition systems have undoubtedly improved the quality of urban crime recordings [4] [5] across

various policing jurisdictions, enabling police practitioners to enhance their understanding of crime dynamics. Fine-grained crime data is particularly useful in hotspot policing, where it is used to identify problematic areas and target appropriate policing responses [6] [7]. However, with improved data quality arise concerns regarding the confidentiality of personal information [8] [9]. The disclosure of personal information from crime data is a serious concern, as it can put individuals at risk of harm, discrimination, or other negative consequences. As such, police agencies take steps to protect the confidentiality of individuals relating to the data. This includes implementing strict data sharing protocols, limiting access to data, and ensuring that any data released is anonymized. Data aggregation is one common technique used to coarsen spatiotemporal data for the purpose of sharing while protecting privacy [10] [11]. However, these techniques can also have negative impacts on the data accuracy [12] [13] [14] [15], data quality [16] [17], and data fitness for purpose [18] [19]. While aggregation (spatial) may serve to reduce biases in analytical outcomes [20], fine-grained raw data sets are often considered more valuable due to their flexibility for manipulation and suitability for a wider range of purposes.

Accessing detailed spatiotemporal crime records presents a set of formidable challenges. Chief among them is the intricacy of data privacy and security [21]. Police is responsible for sensitive and confidential data related to criminal activity, and must ensure that any data sharing is done in compliance with legal requirements and security protocols. Moreover, the fragmentation of data sources across jurisdictions poses a significant hurdle. The diverse methods, formats, and standards of data collection within different geographical boundaries necessitate complex integration efforts, hindering seamless analysis of spatiotemporal patterns. Additionally, resource constraints within law enforcement agencies can impede data quality and accessibility [22] [23]. Many agencies lack the necessary technological infrastructure and expertise to effectively manage and share the complex spatiotemporal data [22]. As a result, inconsistencies and gaps in data reporting can arise. Overcoming these challenges demands collaborative efforts to ensure data security, integration, and accessibility while maintaining the privacy of individuals and the integrity of ongoing investigations.

As an alternative, a synthetic data that models specific aspects of crime dynamics, such as patterns of biases in crime counting [24], spatial concentration of crimes [25] [26], and target selection by offenders [27], can be developed. However, existing studies lack adaptable methodologies and practical tools to replicate real-world datasets or synthetise predefined patterns and interactions among crime occurrences in both spatial and temporal domains. The significance of crime event interactions in crime science cannot be overstated. Numerous crime phenomena, including repeat victimisation [28] [29] [30] [31], crime concentrations [32], and optimal foraging idea [33] [34] emanate from the interplay between crime events in space and time. Consequently, scrutinizing space-time interactions of crime carries both research and operational benefits.

For instance, the recurrent patterns of residential offenders can guide law enforcement in targeting limited police resources. Hence, this paper addresses this methodological challenge by developing "stppSim" tool in R platform in order to allow reproducibility and advancement in other domains.

Specific criminological theories played important roles in the development of "stppSim" tool. These include the rational choice theory (RCT) [35], routine activity theory (RAT) [36], and crime pattern theory (CPT) [37]. In particular, the RAT describes the conditions that have to be met while the offender moves and interacts with the environment. It states that a crime occurs when three elements, namely; motivated offenders, suitable targets, and the absence of capable guardians, converge in space and time: The use of RAT and other related theories for the simulation studies of crime can be found in many existing literature [38] [39] [40].

In order to simulate crime in a virtual environment, two approaches are commonly used, namely: the agent-based modeling (ABM) [41] [42] and microsimulation (MSM) [43] [44]. These techniques operate at the individual (entity) level and rely on assumptions, domain theories, and previous findings. ABM focuses on the interactions between individuals to produce unexpected outcomes. It can simulate complex social systems and model how individuals' actions affect each other and their environment. MSM, on the other hand, focuses on individual stochastic behavior to generate aggregated/dissagregated patterns. In other words, it can simulate the behavior of large populations by modeling the behavior of individual members. [45] demonstrated that ABM and MSM can be integrated to simulate burglary crimes in a heterogeneous environment by combining street network and land use information. This hybrid approach is considered more dynamic than traditional methods and allows for more realistic simulation of crime patterns.

This article introduces an innovative fusion of ABM and MSM techniques to establish a versatile framework for simulating point events across spatial and temporal dimensions. Complementing this framework is an analytical tool named "stppSim", developed within the R programming platform. The primary objective of this study is to elucidate the operational mechanics of the framework, highlight the tool's functionalities, and offer insights into its pivotal outcomes. By harnessing the potential synergy of ABM and MSM, the stppSim tool creatively replicates crime patterns to align with pre-defined specifications. It simulates the stochastic conduct of individual offenders, their engagements with the environment, resulting in the emergence of crime patterns and interactions spanning both spatial and temporal dimensions. In order to ensure a balance between safeguarding the spatiotemporal identities of real individuals and generating valuable data (*i.e.*, accurate records), the simulation commences by anchoring the process at a higher macro (global) level. This involves capturing the overall behaviour, trends, and patterns of the system without delving into the details of individual components. As simulation parameters are gradually dissipated from a broad perspective to a finer granularity, the framework generates

outputs in alignment with the predefined data structure, concurrently upholding location privacy at the detailed level.

This paper is organized as follows: Section 2 presents a detailed overview of the proposed agent-based microsimulation framework. In Section 3, the implementation of the "stppSim" tool is described, emphasizing its key features and functionalities. The application of the tool for generating synthetic spatiotemporal point patterns of crime is described in Section 4. The last section of the paper discussed the significance of the tool in research and practical contexts, while concluding with essential considerations for users and identifying potential areas for future enhancements.

2. Spatio-Temporal Point Pattern Simulation Framework

The proposed simulation framework is aimed at synthesizing crime events marked by the locations and reference times, through artificial offenders (agents) within a specified geographical region and time period. The objective is to ensure that a significant number of events which are relatively close in space are also relatively close in time [46], according to specified spatial and temporal thresholds, hence the space-time interactions between the events.

The framework consists of two main components: Features Calibration and Model Integration, as shown in Figure 1. The Features Calibration component contains two sets of variables: global and individual level variables. These variables are identified as important to crime modeling based on existing theories. The initial values of these variables are set using expert knowledge and research

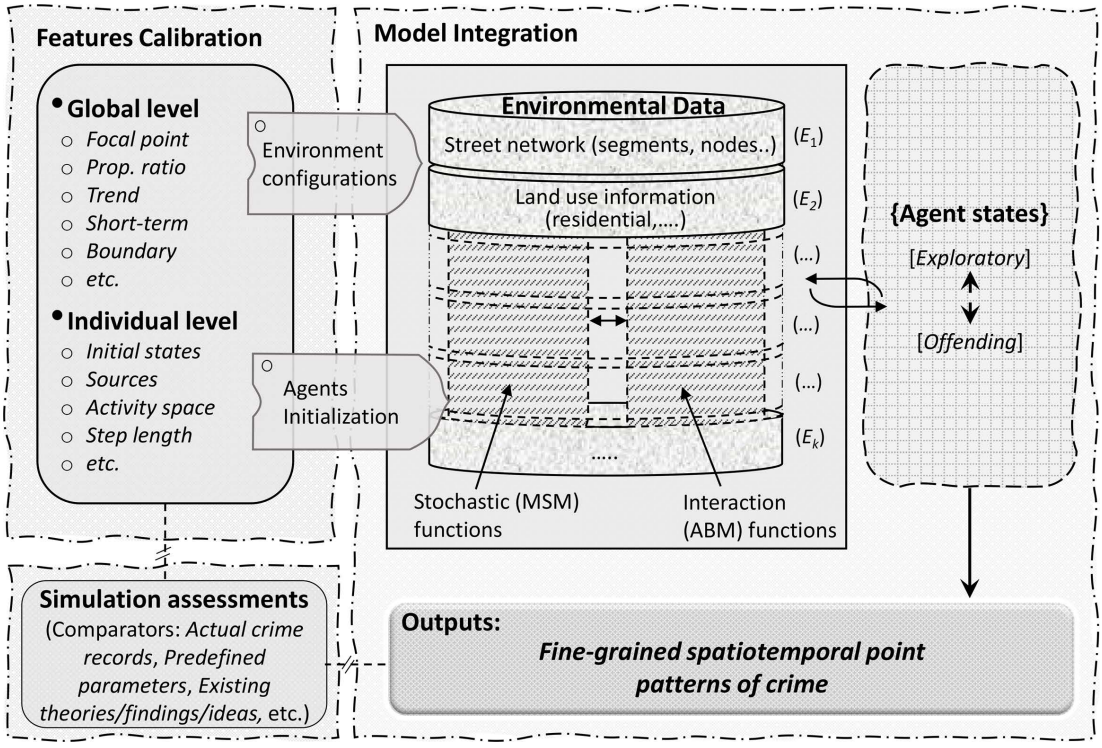

Figure 1. Spatio-temporal point pattern simulation framework.

findings. The global variables are those that affect the overall spatial patterns and trend of crime, such as the spatial proportional ratio [47] and trend direction. On the other hand, individual level variables refer to the characteristics of agents (e.g., offenders) that are embedded in the simulation. These variables may include residences (origins) and speed (step length) of offenders.

The Model Integration component takes the selected variables from the Features Calibration component to initialize and configure the modeling functions (ABMs and MSMs) within the simulation environment. This integration results in the changing of agents' states from "exploratory" to "offending" and vice versa. A crime event is said to occur when an agent assumes an "offending" state. Each component is further described as follows:

2.1. Features Calibration

Using global level variables (see **Table 1**), the framework configures the spatial and temporal properties of the simulation environment. For example, the spatial proportional ratio is a global level variable that controls the spatial concentration of events across the simulation space [47]. Similarly, the trend is a global level variable that determines the long-term direction of the simulated time series. These variables play a critical role in defining the overall characteristics of the simulation environment and ensuring that the synthetic data generated by the simulation is realistic and comparable to the existing data.

The individual level variables connect the global level variables in order to ensure that the simulated data has the desired characteristics at low levels. In other words, the variables control local variations in the simulation, such as the variance in local concentration of events (using "s_band" variable), as well as short-term patterns in the time series. It's important to note that the landscape in which the simulation takes place can be either homogeneous or heterogeneous, with varying levels of restrictions depending on the features included, such as land use or street network.

2.2. Model Integration

In order to initialize and configure the functions that enable agents' movements and interactions across the landscape, the selected variables are integrated together with the environmental features to allow changes in the behaviours (states) of the agents. The process is summarised as pseudocode in **Table 2**.

3. "stppSim" in Practice

The proposed framework is developed into an add-on package "stppSim" to the statistical software R [56]. The utility and the reproducibility are described as follow:

3.1. Implementation

The "stppSim" package is freely available under Open-Source GNU GPL 3

Table 1. Simulation parameters and their descriptions.

Variable	Parameter	Description
Global	Focal point	A point location (x, y) that is considered the focal point of the city. Usually a location within the main central part of the city or an area with the highest concentration of crime opportunities in line with the RAT [36].
	Proportional ratio	The ratio that describes the minimum area to maximum crime concentration in the city, in accordance with law of crime concentration [47].
	Trend	Defines the direction of the time series over time.
	ShortTerm	Specifies the short- to medium-term fluctuations of the time series over time.
	s_range/t_band	A pair of spatial thresholds that defines a geographical region within which point events interact. For guidance on the spatial bandwidth associated with different offender types, see [29] [48] [49] [50] [51].
	Restriction surface	A raster map showing landscape features with different restriction levels (*i.e.*, level of guardianship) [52] [53] depicted by the pixel values. Each feature class will have the same pixel values. At most extreme, the raster may consist of only 0 s (no restrictions) and 1 s (highest restriction). At the other extreme, simulations may be performed without any raster, *i.e.*, on a homogeneous terrain. All intermediate scenarios are also possible. Lower restrictions also imply that an agents can move faster (and so cover more areas and create more interactions), and vice versa.
	Boundary	A shapefile (.shp) object delineating the extent of the landscape. Typically, the.shp object will form the baseline surface on top of which landscape (restrictive) features are stacked. Pixels outside the boundary are by default assigned value 1 s.
Individual level	Initial state	The initial status of an agent. Each agent assumes an '*exploratory*' state at the start of simulation, *i.e.*, no criminal activities is expected. The state may change as a reaction to agent interaction with the environment, driven by the ABM and MSM functions.
	Sources	Defined as a set of x, y coordinates representing the origins of the agents. May be calibrated using the known offender residences [54] [55] or proxy datasets, such as land use patterns or observed crime clustering.
	Spatial threshold	Defines the perception range of an agent at any instant. This is currently defined as a circle around an agent's current location.
	Temporal bin	The temporal unit of analysis. Time intervals within which agents reset (*i.e.*, assumed to re-emerge).
	Step length	Defines the maximum temporal step of an agent. This controls the levels of possible interactions between agents and the environment.

license on the Comprehensive R Archive Network (CRAN) (https://cran.r-project.org/web/packages/stppSim/index.html). The development version and code are available on Github (https://github.com/MAnalytics/stppSim). To install stppSim, open R console (or RStudio) and type: "*install.packages* ('*stppSim*')", then run "library (stppSim)" to load the package.

- Package name: stppSim
- Current version: 1.3.1

Package home page: https://github.com/MAnalytics/stppSim

Table 2. Model integration and state change.

Pseudocode 1:

1. Specify the global temporal parameters over a specified time period, T_k, where k is the number of temporal units in T.
2. Define a specified number of agents, A_Z, in line with the pre-conceived spatial concentration across the landscape,
3. for an agent A_z:
   a) for a temporal bin $k_t \subset T$ and allowed number of time steps, s
      i) Sample the locations around the agent and determine the next agent's location, using the inbuilt (ABM) decision-making process (e.g., preferring low resistance to high resistance location)
      ii) Draw a new state while transiting to the new location $(a_n)_{n\epsilon \mathbf{N}}$ using the in-built stochastic (MSM) function,
      iii) Record agent's current details (*i.e.*, location ID, x, y coordinates, time stamp, state, etc.),
      iv) Return to 4a (i) if n < s.
   b) Commence the next temporal bin, k_t
4. Next agent $(A_i) i \epsilon I$.

- Operating system(s): Platform independent.
- Programming language: R
- Other requirements: R (≥4.1.0)
- Key dependencies: SiMRiv [57]; raster [58]
- License: e.g. GNU GPL v3.0
- Any restrictions to use by non-academics: None。

3.2. Modes of Operation and Assessments

The tool operates in two modes. The first mode allows users to generate complete synthetic data from a sample (source) data, using "*psim_real*" function. The function learns the spatial and temporal properties of the sample data and generates the synthetic dataset accordingly. This method is particularly useful when there is only a sparse or small sample of crime records available. The second mode, using "*psim_artif*" function, generates synthetic data based on pre-defined spatiotemporal characteristics provided by the user, without the need for a sample dataset. This mode is useful when there is no available sample source datasets. In using either of these modes, many of the arguments have been set with default values which are chosen to be suitable for a wide range of scenarios. However, users can re-define any argument to suit their specific research objectives. The detailed instructions and reproducible examples can be found in the package manual and vignette.

The efficacy of the "stppSim" tool can be evaluated through both visual inspection and basic statistical methods. From a visual standpoint, the spatiotemporal patterns can be observed by mapping the distribution of points and track-

ing event trajectories over a given period. Here, using a scatterplot for spatial distribution and a time series plot for temporal patterns would be most appropriate. Conversely, the space-time interactions present in point datasets can be scrutinized with the NearRepeat calculator [59]. This tool determines the statistical significance of proximity of points in both space (within a set spatial range) and time (within a specified temporal frame).

4. Applications

4.1. Replicating Spatio-Temporal Patterns

Besides its ability to generate pre-conceived spatiotemporal point patterns across a research area (via the "*psim_artif*" function), "stppSim" also excels in discerning the spatiotemporal patterns and trends present in a sample source dataset (through the "*psim_real*" function) and subsequently curating new datasets based on those patterns. To illustrate the proficiency of stppSim in this regard, we utilized a randomly chosen subset of residential burglary records from a section of Southwest-side Detroit (Michigan, US) [60]. This subset facilitated the creation of a fully synthesized 1960 events. **Figure 2** offers a visual comparison of the spatial and temporal point arrangements between the source sample (**Figure 2A(i-ii)**), representing 40% (or around 780 records), and the comprehensive original dataset (**Figure 2B(i-ii)**) with its 1960 records. Notably, the sample datasets exhibit a spatial point distribution (SPD) akin to their full dataset counterparts. Moreover, the time series (TS) plots of both groups align considerably, with the overall trends appearing more consistent than their medium-term variations, which, in turn, are more consistent than their short-term variations.

Figure 3A(i) displays the spatial point distribution (SPD) of the synthetic dataset, with its corresponding time series (TS) plot illustrated in **Figure 3A(ii)**. At a glance, the SPD and prominent hotspots closely mirror those of the original datasets, particularly the pronounced hotspots in the southern and western sectors. Nonetheless, some differences are evident, such as the missing cluster in the area's southwest corner within the synthetic data. This omission could be counteracted by leveraging the interactive argument of the function, which previews potential results prior to initiating the simulation. Regions marked as off-limits (like parks or swamps) based on land use data are consistently bypassed in the simulation. For instance, the tract between the south-west corner and the west is designated as a swamp ("resistriction value of 1"). The original dataset's hotspots seem more condensed than their synthetic counterparts, which appear slightly dispersed. A possible remedy could be to use fewer origins during the simulation. Users are encouraged to refer to the package manual for a deeper understanding of how varying parameters can impact the simulation. On a more detailed scale, such as street-level units, there are discernible variances between the source and synthetic datasets. The correlation between the original and synthetic data stands at 0.07, suggesting that pinpoint event locations in the original data don't typically correspond with those in the synthetic set.

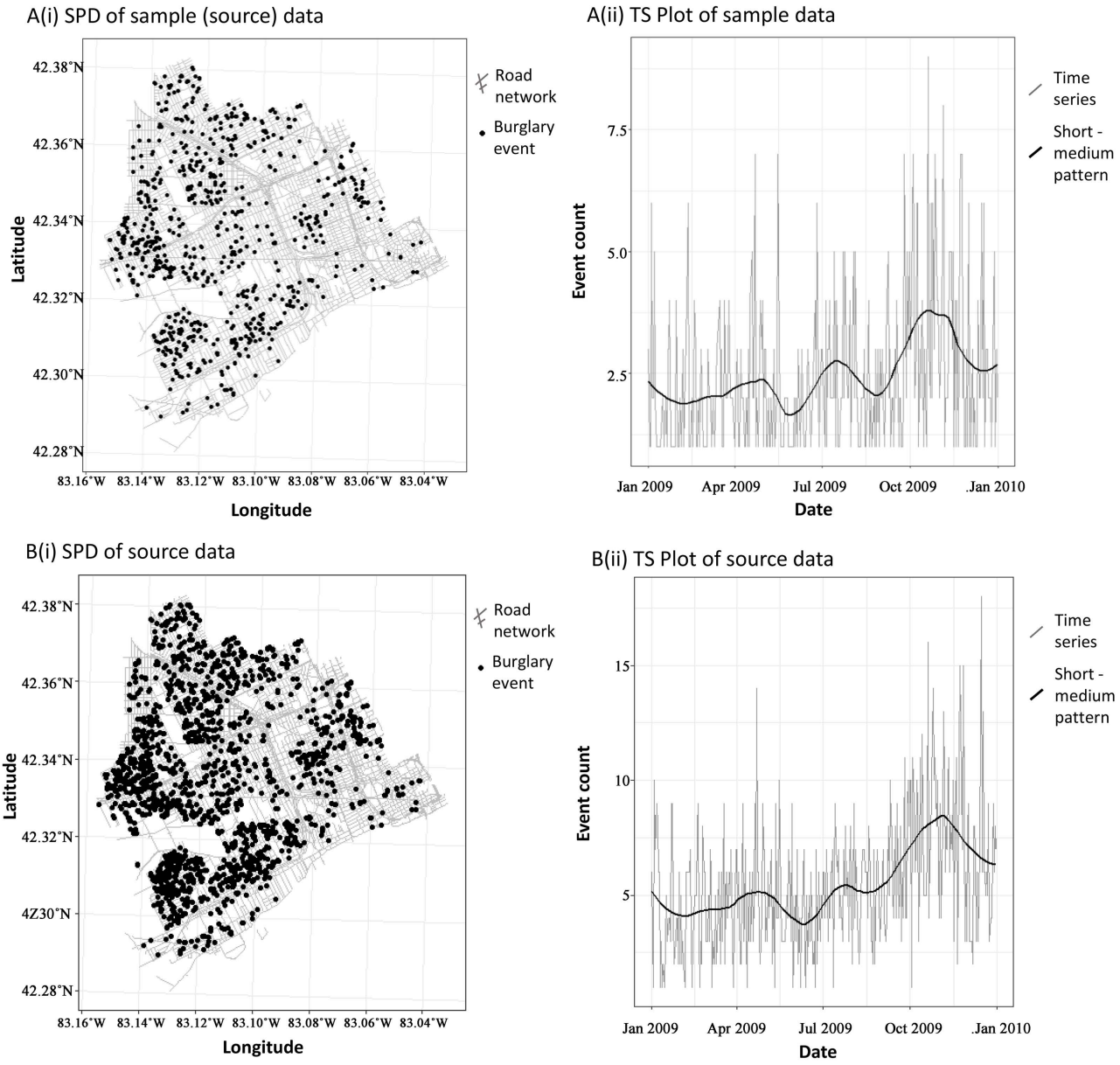

Figure 2. Spatial and temporal pattern of residential burglary of South-west Detroits (Michigan, US, 2009-2010).

In the long term, the most striking resemblance between the two datasets lies in their overarching trend. Both sets, for instance, display a consistent upward trajectory. While the synthetic data showcases more pronounced seasonal spikes compared to the original, their general patterns remain analogous. When observed at a finer temporal resolution, like daily aggregates, the datasets seem more disparate. This distinctiveness in both spatial and temporal detail is essential, ensuring that the precise spatiotemporal locations of individual events in the source datasets remain confidential.

4.2. Simulating Space-Time (ST) Interactions

Utilizing the "psim_real" and "psim_artif" functions, users can respectively

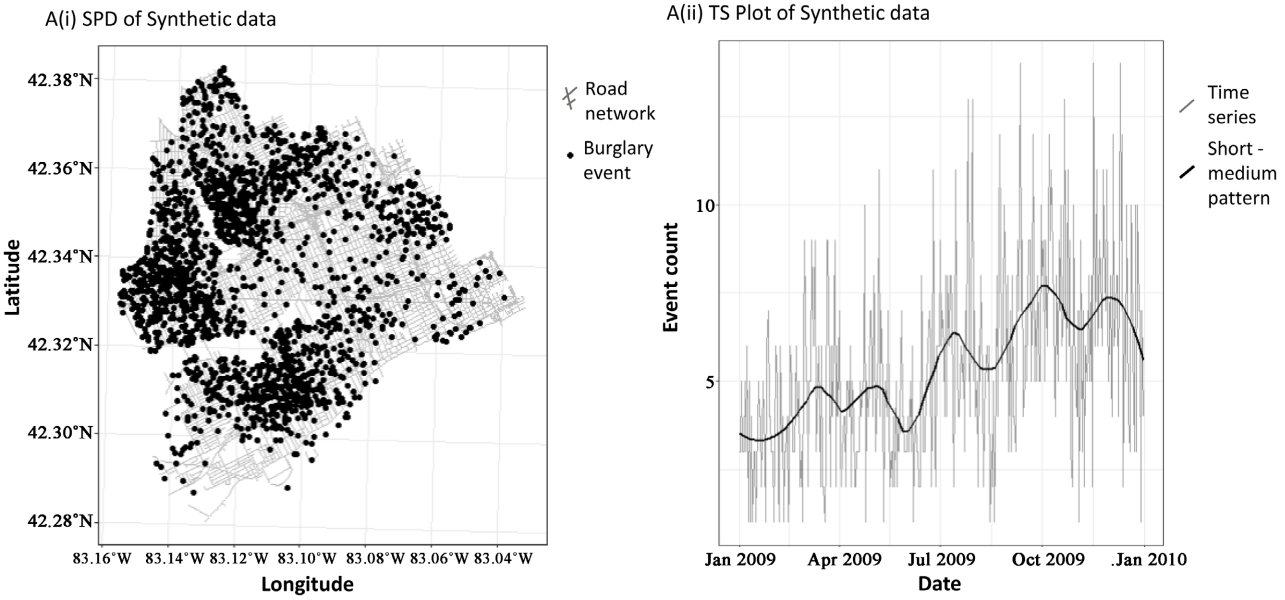

Figure 3. Spatial and temporal pattern of synthetic data.

replicate the space-time interactions found in source dataset and create new synthetic data sets with specified space-time interactions.

1) Emulating Patterns from the Original Dataset

Drawing from the repeat victimization study on residential burglary, a maximum spatial boundary of 600 metres (*i.e.*, s_range = 600) is established. This range is then divided into three equal spatial ranges: (0 - 200 m), (201 - 400 m), and (401 - 600 m), which are labeled as "small", "medium", and "large" spatial bandwidths, respectively. By setting a time span of 30 days with a daily incremental range, Table 3 juxtaposes the outcomes derived from the sample source data (780 records), the full source data (1960 records), and the synthesized data for each bandwidth (1960 records each). The table presents the Knox ratios as per the NearRepeat Calculator, with asterisks denoting statistically significant point interactions for the given space-time bandwidths.

When contrasting the outcomes of the synthetic datasets with both the sample and the full source datasets, it's evident that the package often yields results more akin to the sample datasets than the entire data collection. For instance, within the initial time span (*i.e.*, days 1 - 15), there exist seven overlapping spatiotemporal bandwidths with significant interactions, such as (0 - 200 m) at 6 days, between the synthetic and sample source datasets. In comparison, there are only five shared bandwidths displaying significant interactions, like (201 - 400 m) at 12 days, when matching the synthetic data with the full source datasets.

Transitioning to the latter segment of the time frame (*i.e.*, days 16 - 30), the sample source data shows six concurrent spatiotemporal bandwidths with significant interactions, while the full source data only offers one. Furthermore, the data suggests that interactions within the proximate temporal span (days 1 - 15) are pinpointed with greater precision than those in the extended temporal range (*i.e.*, days 16 - 30). Overall, the outcomes demonstrate the proficiency of stppSim

Table 3. Comparing space-time interactions of source and synthetic datasets.

Dataset	τ (days) / ∂ (m)	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]
Sample source data (780)	(0, 200]	0.73	0.93	**1.91***	1.48	1.37	**1.54***	**1.66***	0.92	**1.64***	1.57	1.72	0.92	1.15	1.53	0.4
	(200, 400]	0.79	1.04	1.12	1.28	1.26	1.14	0.97	0.49	0.87	0.89	0.76	**1.55***	**1.5***	**1.5***	1.13
	(400, 600)	0.86	1.14	1.17	1.14	0.84	1.25	0.92	0.77	0.68	0.95	1.38	1.12	0.96	1.31	0.98
Full source data (1960)	(0, 200]	1.25	1.09	**1.54***	**1.35***	1.24	0.82	**1.39***	1.06	**1.36***	1.21	1.03	0.94	1.27	1.18	1.08
	(200, 400]	1.15	1.17	0.95	1.06	1.07	1.15	1.05	1.17	0.97	1.02	1.13	1.09	**1.43***	**1.23***	1
	(400, 600)	1	1.14	1.12	1.09	1.05	1.07	1.09	0.93	1.03	1.11	1.13	1.05	1.08	0.95	**1.17***
Synthetic data (1960)	(0, 200]	1.03	0.99	**2.11***	0.93	0.73	**1.24***	**1.66***	0.77	**1.96***	0.74	0.9	1.04	1.13	1.22	0.97
	(200, 400]	0.81	0.99	0.95	1	1.05	1	1	1.09	1.02	0.94	0.82	**1.17***	**1.84***	**1.27***	0.99
	(400, 600)	0.99	0.97	1	0.96	1.01	1.01	0.98	0.83	0.9	1.09	**1.19***	0.99	1.04	0.98	0.95
	τ (days) / ∂ (m)	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]
Sample source data (780)	(0, 200]	1.36	1.37	0.2	1.34	1.16	**1.9***	0.59	1.4	0.61	1.73	1.17	0.76	1.38	0.99	1
	(200, 400]	0.99	1.27	1.45	0.64	1.01	0.74	1.42	**1.57***	1.34	**1.55***	1.27	0.42	1.02	1.23	0.95
	(400, 600)	1.3	0.93	0.67	0.98	1.15	1.02	1.1	1.13	1.01	0.66	**1.51***	**1.42***	**1.49***	1.26	**1.69***
Full source data (1960)	(0, 200]	1.2	1.26	**1.32***	1.16	0.87	1.23	**1.34***	**1.34***	0.85	**1.32***	1.11	1.1	**1.47***	1.11	1.21
	(200, 400]	0.98	1.16	1.08	1	0.94	1.01	**1.32***	1.18	**1.28***	1.04	0.91	0.73	1.05	1.1	0.93
	(400, 600)	0.93	1.14	1.02	0.93	1.03	1.09	0.94	1.07	0.88	0.92	1.12	1.11	1	1.09	0.95
Synthetic data (1960)	(0, 200]	1.17	0.96	**1.19***	1.09	0.85	**2.07***	1.09	1.06	**1.47***	1	0.8	0.8	0.66	1.11	0.69
	(200, 400]	0.96	1	1.05	0.95	1.03	0.98	0.85	**1.45***	1	**1.18***	1.05	0.98	0.96	0.92	0.95
	(400, 600)	0.97	0.94	1.03	0.97	0.97	0.98	1.09	0.89	1.05	0.94	**1.45***	1.03	**1.13***	0.88	**1.92***

Signif. codes: $p < 0.001$ "*".

in mirroring spatiotemporal interactions present in source datasets.

2) Simulation of Pre-defined ST Interaction

In the same study region (a segment of Detroit's Southwest side), we generated a synthetic dataset featuring simulated spatiotemporal interactions. Here, three distinct spatial bandwidths were defined: [0 - 100 m], [100 - 200 m], and [200 - 300 m]. Concurrently, four 2-day interval temporal bandwidths were specified: 4 - 5, 13 - 14, 21 - 22, and 28 - 29 days. As a result, twelve individual synthetic datasets were formulated. Each dataset encapsulates point interactions as characterized by a distinct combination of spatiotemporal bandwidths that mirror the actual bandwidths.

For every synthetic dataset, the NearRepeat calculator is utilized to evaluate all potential combinations of spatiotemporal bandwidths (hereafter referred to as test bandwidths) juxtaposed against the actual bandwidths. Table 4 showcases these findings. It's worth noting that the diagonally aligned cells with statistical significance denote that the spatiotemporal interactions at the pertinent real spatial

Table 4. Detection of space-time interactions in synthetic data.

Test bandwidths		*Real bandwidths*											
τ (days)		[4, 5]			[13, 14]			[21, 22]			[27, 28]		
	∂ (m)	(0, 100]	(100, 200]	(200, 300)	(0, 100]	(100, 200]	(200, 300)	(0, 100]	(100, 200]	(200, 300)	(0, 100]	(100, 200]	(200, 300)
[4, 5]	(0, 100]	**3.72***	1.08	0.33	0.93	0.89	0.58	0.68	0.94	0.97	0.58	0.39	0.55
	(100, 200]	**2.05***	**2.05***	0.73	1.03	0.86	0.58	0.85	0.78	0.25	0.56	0.8	0.88
	(200, 300)	0.85	**1.67***	**1.52***	1	0.93	0.64	**1.27***	0.74	0.14	0.65	0.91	**1.2***
[13, 14]	(0, 100]	0.97	0.76	0.76	**2.82***	1.1	0.48	0.26	0.71	1.09	0.54	0.75	0.75
	(100, 200]	0.13	0.94	0.72	**1.66***	**1.82***	0.82	0.44	0.73	0.71	0.64	0.76	**1.14***
	(200, 300)	0.99	1.01	0.81	0.77	**1.79***	**1.77***	0.58	0.83	1.05	0.22	0.74	0.93
[21, 22]	(0, 100]	0.69	0.8	0.28	0.73	0.71	0.49	**3.59***	0.7	0.92	0.85	0.23	0.92
	(100, 200]	0.81	0.8	**1.2***	0.89	1.02	0.86	**1.76***	**1.82***	0.88	0.78	**1.18***	1.01
	(200, 300)	**1.12***	**1.13***	**1.26***	0.93	0.93	1.05	0.9	**1.3***	**1.36***	1.02	0.36	0.86
[27, 28]	(0, 100]	0.55	0.81	0.37	**3.59***	0.81	**2.37***	0.83	0.69	0.65	**3.41***	1.11	0.93
	(100, 200]	0.64	0.93	1	**2.55***	**1.95***	**1.9***	1.03	1.02	0.86	**2.02***	**2.24***	0.97
	(200, 300)	0.95	0.93	0.17	**1.29***	**1.88***	**1.54***	0.65	1.02	0.92	0.71	**1.84***	**1.76***

Signif. codes: $p < 0.001$ "*".

and temporal bandwidths were simulated with accuracy.

Each of the twelve synthetic datasets effectively mirrored the intended point interactions, as evidenced by the significant findings in the diagonal cells. This reaffirms the tool's prowess in precisely emulating spatiotemporal interactions.

It's also noteworthy to mention the presence of significant results in off-diagonal cells. Such findings can be credited to the compounded effects of the bandwidths specified. Essentially, a set spatial or temporal bandwidth can inadvertently catalyze the manifestation of point interactions across broader spatial or temporal bandwidths, especially if these larger bandwidths are direct multiples of the original ones. A clear illustration of this phenomenon can be observed in cells adjacent to the diagonal with significant results. For instance, the 0 - 100 m spatial bandwidth could be perceived as an inherent component of its larger counterparts, namely; 100 - 200 m and 200 - 300 m. Furthermore, beyond just the spatial dimensions, the notable groupings of statistically significant cells within the temporal bandwidths of 21 - 22 days and 27 - 28 days can be traced back to the cumulative influences of the 4 - 5 days and 13 - 14 days temporal bandwidths, respectively.

5. Discussion and Conclusions

Given the limited availability of detailed crime records, the "stppSim" package

serves as a valuable data resource for both research and educational purposes. It provides an alternative data source that can facilitates in-depth examination of crime dynamics in space and time, leading to potential policy and operational implications. The package is conveniently accessible on the CRAN platform, allowing users to freely download, reuse, redistribute, and explore its applications in various domains.

The field of criminology recognizes the significant value of examining the space-time interaction of crime events in various contexts. In the analysis of recurring residential burglaries, such analysis can aid in identifying individuals and locations that face a disproportionate risk of victimization [29] [48]. Researchers are often interested not only in the "same repeat" victims, referring to individuals or locations that experience multiple crimes within a short period after the initial incident, but also in the concept of "near repeat" victims. These near repeat victims are nearby individuals or locations that become victimized shortly after the initial crime occurs. The stppSim package provides opportunities for simulating or exploring different scenarios of repeat and near-repeat victimization within a specific geographical area. As demonstrated in this paper using a section of South-west Detroit as an example, it becomes possible to identify the spatial and/or temporal signatures associated with a particular area [47] [61] [62].

The analysis of spatio-temporal point interaction extends beyond criminology and finds applications in various research domains, such as earthquakes, ecology, epidemiology, and more. In these fields, the identification of relationships between events and their evolution over space and time holds significant importance. Researchers seek to understand the underlying phenomenon by studying spatio-temporal event interactions, clustering or regularity patterns, and distances that provide insights into these interactions. For instance, in ecological studies at the community level, the analysis focuses on examining interactions of competition and facilitation among trees as a primary objective.

There are several important considerations for potential users of the stppSim package. Firstly, it should be noted that the synthetic point events generated by the package are inherently geomasked at a fine-grained level. This is done to preserve the sensitivity of any source data used. Secondly, the simulation functions in the package incorporate specific random elements, which means that two synthetic datasets generated with the same simulation parameters may not be identical. However, an "interactive" argument embedded in the functions can be used to preview the spatio-temporal models before continuing the actual simulation. Thirdly, it's important to recognize that the properties of the synthetic data may be biased towards the characteristics of the sample dataset provided, rather than accurately representing the entire population of the actual data. Therefore, synthetic data should not be considered a replacement for real or source data. Any modeling or inference conducted on synthetic data carries additional risks. The author of the package suggests that synthetic data, when used in a research context, can help expedite the research process, but it is crucial that

any final data intended for real-world applications be evaluated and fine-tuned using the actual data if necessary. Lastly, it should be noted that the current version of the stppSim package is computationally intensive, particularly when using the "psim_real" function. On a standard office PC with an Intel Core i7-7500CPU and 16.0 GB RAM, it takes approximately 30 minutes to complete. However, the "psim_artif" function allows for the generation of synthetic data within a relatively short period, such as around 5 minutes. Future work on the package will prioritize improving computational efficiency by incorporating parallel processing functions. Additionally, upcoming versions of the package will include the ability to simulate other relevant nominal information, such as age, gender, occupation, and so on, of the objects under study. The author of the package encourages users to provide suggestions, feedback, bug reports, and explore opportunities for collaborations to further enhance its capabilities.

In summary, while stppSim offers valuable synthetic data generation capabilities, users should be aware of the geomasking, inherent randomness, and potential biases of the synthetic data. It is essential to exercise caution and verify results with real data when applying the findings to real-world scenarios.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Crawford, A. and Evans, K. (2017) Crime Prevention and Community Safety. In: Liebling, A., Maruna, S. and McAra, L., Eds., *The Oxford Handbook of Criminology*, Oxford University Press, Oxford, 797-824. https://doi.org/10.1093/he/9780198719441.003.0036

[2] Andresen, M.A. and Malleson, N. (2015) Intra-Week Spatial-Temporal Patterns of Crime. *Crime Science*, **4**, Article No. 12. https://doi.org/10.1186/s40163-015-0024-7

[3] Johnson, S.D. (2010) A Brief History of the Analysis of Crime Concentration. *European Journal of Applied Mathematics*, **21**, 349-370. https://doi.org/10.1017/S0956792510000082

[4] Bakardjiev, D.K. (2015) Officer Body-Worn Cameras-Capturing Objective Evidence with Quality Technology and Focused Policies. *Jurimetrics*, **56**, 79-112.

[5] O'Connor, C.D., Ng, J., Hill, D. and Frederick T. (2022) Thinking about Police Data: Analysts' Perceptions of Data Quality in Canadian Policing. *The Police Journal*, **95**, 637-656. https://doi.org/10.1177/0032258X211021461

[6] Telep, C.W., Mitchell, R.J. and Weisburd, D. (2014) How Much Time Should the Police Spend at Crime Hot Spots? Answers from a Police Agency Directed Randomized Field Trial in Sacramento, California. *Justice Quarterly*, **31**, 905-933. https://doi.org/10.1080/07418825.2012.710645

[7] Leigh, J., Dunnett, S. and Jackson, L. (2019) Predictive Police Patrolling to Target Hotspots and Cover Response Demand. *Annals of Operations Research*, **283**, 395-410. https://doi.org/10.1007/s10479-017-2528-x

[8] Lin, R. (2015) Police Body Worn Cameras and Privacy: Retaining Benefits While Reducing Public Concerns. *Duke Law & Technology Review*, **14**, 346.

[9] Poullet, Y. (2004) The Fight against Crime and/or the Protection of Privacy: A Thorny Debate! *International Review of Law, Computers & Technology*, **18**, 251-273. https://doi.org/10.1080/1360086042000223535

[10] Sherman, J.E. and Fetters, T.L. (2007) Confidentiality Concerns with Mapping Survey Data in Reproductive Health Research. *Studies in Family Planning*, **38**, 309-321. https://doi.org/10.1111/j.1728-4465.2007.00143.x

[11] Wiggins, L. (2002) Using Geographic Information Systems Technology in the Collection, Analysis, and Presentation of Cancer Registry Data: A Handbook of Basic Practices. North American Association of Central Cancer Registries, Springfield IL, 33-34.

[12] Zhang, Z., Zhang, H., Zhao, L., Chen, T., Arik, S.Ö. and Pfister, T. (2022) Nested Hierarchical Transformer: Towards Accurate, Data-Efficient and Interpretable Visual Understanding. *Proceedings of the AAAI Conference on Artificial Intelligence.* **36**, 3417-3425. https://doi.org/10.1609/aaai.v36i3.20252

[13] Jeffery, C., Ozonoff, A. and Pagano, M. (2014) The Effect of Spatial Aggregation on Performance When Mapping a Risk of Disease. *International Journal of Health Geographics*, **13**, Article No. 9. https://doi.org/10.1186/1476-072X-13-9

[14] Hornberger, Z.T., Cox, B.A. and Hill, R.R. (2019) Effects of Aggregation Methodology on Uncertain Spatiotemporal Data. arXiv: 1910.05125.

[15] Geiger, E.F., Heron, S.F., Hernández, W.J., *et al.* (2021) Optimal Spatiotemporal Scales to Aggregate Satellite Ocean Color Data for Nearshore Reefs and Tropical Coastal Waters: Two Case Studies. *Frontiers in Marine Science*, **8**, Article 643302. https://doi.org/10.3389/fmars.2021.643302

[16] Karlson, R.H., Cornell, H.V. and Hughes, T.P. (2007) Aggregation Influences Coral Species Richness at Multiple Spatial Scales. *Ecology*, **88**, 170-177. https://doi.org/10.1890/0012-9658(2007)88[170:AICSRA]2.0.CO;2

[17] McGlinn, D.J., Engel, T., Blowes, S.A., *et al.* (2021) A Multiscale Framework for Disentangling the Roles of Evenness, Density, and Aggregation on Diversity Gradients. *Ecology*, **102**, e03233. https://doi.org/10.1002/ecy.3233

[18] Liaw, S.T., Taggart, J., Dennis, S. and Yeo, A. (2011) Data Quality and Fitness for Purpose of Routinely Collected Data—A General Practice Case Study from an Electronic Practice-Based Research Network (ePBRN). *AMIA Annual Symposium Proceedings*, **2011**, 785-794.

[19] Errington, A., Einbeck, J., Cumming, J., Rössler, U. and Endesfelder, D. (2021) The Effect of Data Aggregation on Dispersion Estimates in Count Data Models. *The International Journal of Biostatistics*, **18**, 183-202. https://doi.org/10.1515/ijb-2020-0079

[20] Buil-Gil, D., Medina, J. and Shlomo, N. (2021) Measuring the Dark Figure of Crime in Geographic Areas: Small Area Estimation from the Crime Survey for England and Wales. *The British Journal of Criminology*, **1**, 364-388. https://doi.org/10.1093/bjc/azaa067

[21] Fox, J.C. and Lundman, R.J. (1974) Problems and Strategies in Gaining Research Access in Police Organizations. *Criminology*, **12**, 52-69. https://doi.org/10.1111/j.1745-9125.1974.tb00620.x

[22] Gottschalk, P. (2006) Knowledge Management Systems in Law Enforcement: Technologies and Techniques. IGI Global, Hershey. https://doi.org/10.4018/978-1-59904-307-4

[23] Ashby, D.I., Irving, B.L. and Longley, P.A. (2007) Police Reform and the New Public Management Paradigm: Matching Technology to the Rhetoric. *Environment and*

Planning C: *Government and Policy*, **25**, 159-175. https://doi.org/10.1068/c0556

[24] Brunton-Smith, I., Buil-Gil, D., Pina-Sánchez, J., Cernat, A. and Moretti, A. (2023) Using Synthetic Crime Data to Understand Patterns of Police Under-Counting at the Local Level. https://www.crimrxiv.com/pub/2j7s2j6z

[25] Devia, N. and Weber, R. (2013) Generating Crime Data Using Agent-Based Simulation. *Computers, Environment and Urban Systems*, **42**, 26-41. https://doi.org/10.1016/j.compenvurbsys.2013.09.001

[26] Rosés, R., Kadar, C. and Malleson, N. (2021) A Data-Driven Agent-Based Simulation to Predict Crime Patterns in an Urban Environment. *Computers, Environment and Urban Systems*, **89**, Article ID: 101660. https://doi.org/10.1016/j.compenvurbsys.2021.101660

[27] Malleson, N., Heppenstall, A., See, L. and Evans, A. (2013) Using an Agent-Based Crime Simulation to Predict the Effects of Urban Regeneration on Individual Household Burglary Risk. *Environment and Planning B*: *Planning and Design*, **40**, 405-426. https://doi.org/10.1068/b38057

[28] Farrell, G. and Pease, K. (2001) Repeat Victimization. Criminal Justice Press, Monsey, New York.

[29] Bowers, K.J. and Johnson, S.D. (2004) Who Commits Near Repeats? A Test of the Boost Explanation. *Western Criminology Review*, **5**, 12-24.

[30] Farrell, G. (1995) Preventing Repeat Victimization. *Crime and Justice*, **19**, 469-534. https://doi.org/10.1086/449236

[31] Grove, L.E., Farrell, G., Farrington, D.P. and Johnson, S.D. (2012) Preventing Repeat Victimization: A Systematic Review. Brottsförebyggande rådet/The Swedish National Council for Crime Prevention, Stockholm.

[32] Weisburd, D. and Braga, A.A. (2006) Advocate Hot Spots Policing as a Model for Police Innovation. In: Weisburd, D. and Braga, A., Eds., *Police Innovation*: *Contrasting Perspectives*, Cambridge University Press, Cambridge, 225-244. https://doi.org/10.1017/CBO9780511489334.012

[33] Halford, E. (2023) Linking Foraging Domestic Burglary: An Analysis of Crimes Committed within Police-Identified Optimal Forager Patches. *Journal of Police and Criminal Psychology*, **38**, 127-140. https://doi.org/10.1007/s11896-022-09497-8

[34] Fielding, M. and Jones, V. (2012) 'Disrupting the Optimal Forager': Predictive Risk Mapping and Domestic Burglary Reduction in Trafford, Greater Manchester. *International Journal of Police Science & Management*, **14**, 30-41. https://doi.org/10.1350/ijps.2012.14.1.260

[35] Clarke, R.V. and Cornish, D.B. (1985) Modeling Offenders' Decisions: A Framework for Research and Policy. *Crime and Justice*, **6**, 147-185. https://doi.org/10.1086/449106

[36] Cohen, L.E. and Felson, M. (1979) Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review*, **44**, 588-608. https://doi.org/10.2307/2094589

[37] Brantingham, P.L. and Brantingham, P.J. (1993) Environment, Routine and Situation: Toward a Pattern Theory of Crime. *Advances in Criminological Theory*, ***5***, 259-294. https://doi.org/10.4324/9781315128788-12

[38] Groff, E.R. (2007) Simulation for Theory Testing and Experimentation: An Example Using Routine Activity Theory and Street Robbery. *Journal of Quantitative Criminology*, **23**, 75-103. https://doi.org/10.1007/s10940-006-9021-z

[39] Groff, E. (2008) Simulating Crime to Inform Theory and Practice. In: Chainey, S.

and Tompson, L., Eds., *Crime Mapping Case Studies: Practice and Research*, Wiley, New York, 133. https://doi.org/10.1002/9780470987193.ch16

[40] Malleson, N., Heppenstall, A. and Crooks, A. (2018) Place-Based Simulation Modeling: Agent-Based Modeling and Virtual Environments. In: *Oxford Research Encyclopedia of Criminology and Criminal Justice*, Oxford University Press, Oxford University, UK. https://doi.org/10.1093/acrefore/9780190264079.013.319

[41] Gerritsen, C. and Elffers, H. (2020) Agent-Based Modelling for Criminological Theory Testing and Development. Routledge, London. https://doi.org/10.4324/9780429277177

[42] Groff, E.R., Johnson, S.D. and Thornton, A. (2019) State of the Art in Agent-Based Modeling of Urban Crime: An Overview. *Journal of Quantitative Criminology*, **35**, 155-193. https://doi.org/10.1007/s10940-018-9376-y

[43] Rephann, T.J. and Öhman, M. (1999) Building a Microsimulation Model for Crime in Sweden: Issues and Applications.

[44] Adepeju, M.O. and Evans, A. (2018) A Dynamic Microsimulation Framework for Generating Synthetic Spatiotemporal Crime Patterns. GISRUK Proceedings, Leeds.

[45] Malleson, N. and Birkin, M. (2012) Analysis of Crime Patterns through the Integration of an Agent-Based Model and a Population Microsimulation. *Computers, Environment and Urban Systems*, **36**, 551-561. https://doi.org/10.1016/j.compenvurbsys.2012.04.003

[46] Diggle, P.J., Chetwynd, A.G., Häggkvist, R. and Morris, S.E. (1995) Second-Order Analysis of Space-Time Clustering. *Statistical Methods in Medical Research*, **4**, 124-136. https://doi.org/10.1177/096228029500400203

[47] Weisburd, D. (2015) The Law of Crime Concentration and the Criminology of Place. *Criminology*, **53**, 133-157. https://doi.org/10.1111/1745-9125.12070

[48] Bowers, K.J. and Johnson, S.D. (2005) Domestic Burglary Repeats and Space-Time Clusters: The Dimensions of Risk. *European Journal of Criminology*, **2**, 67-92. https://doi.org/10.1177/1477370805048631

[49] Townsley, M., Homel, R. and Chaseling, J. (2003) Infectious Burglaries. A Test of the Near Repeat Hypothesis. *British Journal of Criminology*, **43**, 615-633. https://doi.org/10.1093/bjc/43.3.615

[50] Johnson, S.D., Davies, T., Murray, A., Ditta, P., Belur, J. and Bowers, K. (2017) Evaluation of Operation Swordfish: A Near-Repeat Target-Hardening Strategy. *Journal of Experimental Criminology*, **13**, 505-525. https://doi.org/10.1007/s11292-017-9301-7

[51] Bediroglu, G., Bediroglu, S., Colak, H.E. and Yomralioglu, T. (2018) A Crime Prevention System in Spatiotemporal Principles with Repeat, Near-Repeat Analysis and Crime Density Mapping: Case Study Turkey, Trabzon. *Crime & Delinquency*, **64**, 1820-1835. https://doi.org/10.1177/0011128717750391

[52] Felson, M. (1987) Routine Activities and Crime Prevention in the Developing Metropolis. *Criminology*, **25**, 911-932. https://doi.org/10.1111/j.1745-9125.1987.tb00825.x

[53] Leclerc, B. and Reynald, D. (2017) When Scripts and Guardianship Unite: A Script Model to Facilitate Intervention of Capable Guardians in Public Settings. *Security Journal*, **30**, 793-806. https://doi.org/10.1057/sj.2015.8

[54] Langton, S.H. (2019) Offender Residential Concentrations: A Longitudinal Study in Birmingham, England.

[55] Savage, J. and Windsor, C. (2018) Sex Offender Residence Restrictions and Sex

Crimes against Children: A Comprehensive Review. *Aggression and Violent Behavior*, **43**, 13-25. https://doi.org/10.1016/j.avb.2018.08.002

[56] R Core Team (2022) A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/

[57] Quaglietta, L. and Porto, M. (2019) SiMRiv: An R Package for Mechanistic Simulation of Individual, Spatially-Explicit Multistate Movements in Rivers, Heterogeneous and Homogeneous Spaces Incorporating Landscape Bias. *Movement Ecology*, **7**, Article No. 11. https://doi.org/10.1186/s40462-019-0154-8

[58] Hijmans, R.J., Van Etten, J., Cheng, J., Mattiuzzi, M., *et al.* (2015) Package 'Raster'. *R Package*, **734**, 473.

[59] Steenbeek, W. (2018) Near Repeat. R Package Version 0.1.1. 2018. https://github.com/wsteenbeek/NearRepeat

[60] Ashby, M.P. (2019) Studying Crime and Place with the Crime Open Database: Social and Behavioural Sciences. *Research Data Journal for the Humanities and Social Sciences*, **4**, 65-80. https://doi.org/10.1163/24523666-00401007

[61] Ratcliffe, J.H. and McCullagh, M.J. (1998) Aoristic Crime Analysis. *International Journal of Geographical Information Science*, **12**, 751-764. https://doi.org/10.1080/136588198241644

[62] Ratcliffe, J.H. (2002) Aoristic Signatures and the Spatio-Temporal Analysis of High Volume Crime Patterns. *Journal of Quantitative Criminology*, **18**, 23-43. https://doi.org/10.1023/A:1013240828824