June 2022

Three Dimensional Spatio-Temporal Cluster Analysis of SARS-CoV-2 Infections

Keith W. Allison
University of Massachusetts Amherst

THREE DIMENSIONAL SPATIO-TEMPORAL CLUSTER ANALYSIS OF SARS-COV-2 INFECTIONS

A Thesis Presented

by

KEITH W. ALLISON

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

May 2022

Department of Biostatistics and Epidemiology

THREE DIMENSIONAL SPATIO-TEMPORAL CLUSTER ANALYSIS OF SARS-COV-2 INFECTIONS

A Thesis Presented

by

KEITH W. ALLISON

Approved as to style and content by:

Andrew Lover, Chair

Laura B. Balzer, Member

John Staudenmayer, Member

Lisa Chasan-Taber, Department Chair
Department of Biostatistics and Epidemiology

ACKNOWLEDGMENTS

Thank you to Andrew Lover for giving me excellent guidance and a fine pineapple plant.

ABSTRACT

THREE DIMENSIONAL SPATIO-TEMPORAL CLUSTER ANALYSIS OF SARS-COV-2 INFECTIONS

MAY 2022

KEITH W. ALLISON

B.Sc., NORTH CAROLINA STATE UNIVERSITY

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Andrew Lover

The COVID-19 pandemic has heightened the need for fine-scale analysis of the clustering of cases of infectious disease in order to better understand and prevent the localized spread of infection. The students living on the University of Massachusetts, Amherst campus provided a unique opportunity to do so, due to frequent mandatory testing during the 2020-2021 academic year, and dense living conditions. The South-West dormitory area is of particular interest due to its extremely high population density, housing around half of students living on campus during normal conditions. Using data gathered by the Public Health Promotion Center (PHPC), we analyzed the clustering of SARS-CoV 2 cases in three-dimensional space as well as time within and between the three tallest occupied buildings in the Southwest dormitory area, John Quincy Adams, Kennedy, and Coolidge. We used the SaTScan program and its Space-Time Permutation Model, which searches for areas with a greater than expected number of cases. Analysis was done at various levels of spacial detail. Additionally,

this analysis was compared to the purely temporal surveillance method, CDC's Early Aberration Reporting System (EARS). Analysis with SaTScan at the room and floor level showed multiple significant clusters within the Coolidge dormitory building. Floor-level analysis was found to be as sensitive as and less burdensome than room-level analysis. We recommend using scan statistics in conjunction with other methods such as purely temporal scans and wastewater analysis to detect and respond to outbreaks on campus.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 COVID-19

The first cases of novel coronavirus disease 2019 (COVID-19) appeared in Wuhan City, China in December of 2019 (Huang et al). This disease is caused by the respiratory virus SARS-CoV-2 which is believed to have transferred from zoonotic sources, most likely from human-animal interactions in live animal markets (Li et al 2021). This virus quickly spread globally, with the WHO declaring a PHEIC (Public Health Emergency of International Concern) in January 2020 (WHO statement, Jan 2020) and a pandemic in March 2020 (WHO media briefing). Public health organizations rapidly began gathering, analyzing, and disseminating data about the spread of SARS-CoV-2. In the United States, major efforts have been made to track and forecast the spread of the virus, including using spatial, temporal, and spatial-temporal methods to detect potential clusters of cases. SARS-CoV-2 infections are an excellent testing-ground for these methods due to massive-scale testing, high-resolution data, and well-understood natural history of infections across the subsequent pandemic waves.

1.2 University of Massachusetts, Amherst Pandemic Response

In response to the pandemic, the University of Massachusetts, Amherst suspended all in-person classes on March 13th, 2020, and on March 20th, 2020 finalized plans to relocate all remaining students living on campus. The campus continued to be empty of students for the remainder of the academic year. The University allowed students

to remain on campus and attend classes during the 2020-2021 academic year, with measures in place to mitigate the spread of the virus. Masking was required in all indoor areas, and for a period of time outdoors as well. In-class learning was minimal (essential lab-based courses only) and most learning was done remotely. Students living on-campus as well as students who did not live on campus but attended in-person classes were required to submit to frequent (weekly, or bi-weekly) testing. Those students who tested positive were moved to isolation residences and thoroughly interviewed by workers and nurses from the University's Public Health Promotion Center (PHPC). This subsequent contact tracing also recorded reported symptoms, and geographic locations. These interview data were captured into a central HIPAA-compliant database for analysis.

1.3 Outbreak Detection

Many outbreak detection algorithms are purely temporal and are the most widely used in public health surveillance (Kulldorff et al 2005). However, spatial-temporal methods are increasingly used for epidemiological surveillance and research. There are multiple scenarios for using these expanded methods. For example, Kan et al used spatial-temporal methods to identify high risk areas in Hong Kong with elevated numbers of reported COVID-19 infections. They suggest that their research can inform policy, lead to more targeted prevention strategies, and impact the behavior of the public. Dellicour et al also used spatial-temporal methods to determine factors associated with COVID- 19 hospitalization to predict future hospitalizations. Ling et al studied dengue fever in Malaysia at both the street address and sub-district (county) level and explored the effects of scale on analysis. This work suggests that incorporation of spatial and temporal elements of dengue must be carefully considered to create efficient control measures. Finally, a large field exists developing "optimal" methods for outbreak detection using spatial-temporal methods. Deeper understanding of the

spatial-temporal nature of outbreaks has the potential to allow for more efficient and effective responses in the future.

1.4 Rationale for Retrospective Analysis

Retrospective analysis of outbreak infection data is useful for optimizing outbreak detection systems currently in place, exploring alternative detection methods, and identifying areas of concern or previously undetected clusters of infection. An outbreak detection system ideally should identify all outbreaks at a given threshold, without giving any false positives. Of course, this requires adjustment and testing of algorithm parameters. Retrospective analysis provides a suitable environment for such testing, as sensitivity can be adjusted to detect outbreaks that were identified by active surveillance or considered clinically significant and ignore those that would not merit a response. Care must be taken to avoid reducing the sensitivity to the point where important anomalies are ignored. New methods of outbreak or cluster detection can also be tested with retrospective datasets. These methods can be tested for effectiveness and ease of use before being implemented for prospective detection.

The data collected by the PHPC had a high spatial resolution compared to other spatial data collected for outbreak detection purposes, including both floor and room numbers. Cluster detection methods generally use data at larger scales, like county or zip code, or neighborhood. Well-georeferenced data in high-density living environments are important epidemiological tools, and due to institutional record keeping often provide critical epidemiological insights that would not be possible using routine household level data. Even event detection done at the street or address level lacks information about floor and room location in space. To have data at this resolution requires specific workflows and data collection as well as building floor plans and complex and time-consuming geolocation that cannot currently be done by address geolocation APIs. This high-resolution spatial data may allow us to better pinpoint

problematic areas at the floor or room level rather than at the address level, and to quantify patterns of transmission in novel ways. This would allow the PHPC to more efficiently allocate limited resources to contain the spread of infections within large dormitories or other congregate living environments. Finally, incorporation of wastewater sampling techniques to identify dormitories with elevated viral loads and using cluster detection methods at a room level could be an effective and precise way to isolate areas of increased infection.

1.5 Goals

The spread of infectious diseases in both space and time has been extensively studied, especially in the context of emerging diseases like COVID-19. The Space-time Scan Statistic (STSS), developed by Kulldorf and collaborators, combines both the spatial and temporal elements of disease analysis to detect clusters, and has been used both prospectively, to predict new clusters of COVID-19, and retrospectively, to analyze previously collected data for clusters. These studies are often done at a ZIP code or county-level. Such is the case in Rosillo et al. (2021) and Xu and Beard (2020). However, STSS is capable of cluster detection at finer spatial scales. This could be at the address level or even down to the room level, in the case of apartment buildings or dormitories. This level of resolution can help detect the localized spread of disease and has been done by Abboud et al to study the spread of *Klebsiella pneumoniae* in a single, high complexity hospital. Our goal is to assess the implementation of STSS at this high-resolution, room-specific, level and determine if significant clustering occurs within dormitory buildings, and to assess sensitivity in comparison to coarser spatial scales.

CHAPTER 2

METHODS

2.1 Data acquisition and processing

We accessed the data collected by the UMass Public Health Promotion Center (PHPC) about SARS-CoV-2 infections occurring in the John Quincy Adams, Kennedy, and Coolidge dormitory buildings during the period of time from February 1st, 2021, to May 15th 2021. This time period coincided with the 2021 spring semester. These data included for each individual the date of infected sample collection, the date of symptom onset, the dormitory building, gender, floor number, and room number. The date of infection was chosen from the date of sample collection or reported date of symptom onset, whichever is earlier. Figure 2.1 shows the number of cases by date for each of the three dorms. All data were collected as part of routine PHPC public health response, which was reviewed by UMass IRB, and determined to not be human subjects research (HRPO reference number: 20-258).

We used QGIS (https://qgis.org/en/site/, version 3.22.3) to overlay floor plans obtained from the UMass residential services website on building latitude-longitude, referenced using Google earth. We determined the location of each room by geolocating its center point. The elevation of each room was approximated using the total height of the buildings, and assuming evenly-spaced floors. Together, these data allowed us to locate the relative position of reported each case in three-dimensional space. The distances between cases were determined in a Euclidean fashion; that is, in direct lines between cases in space.

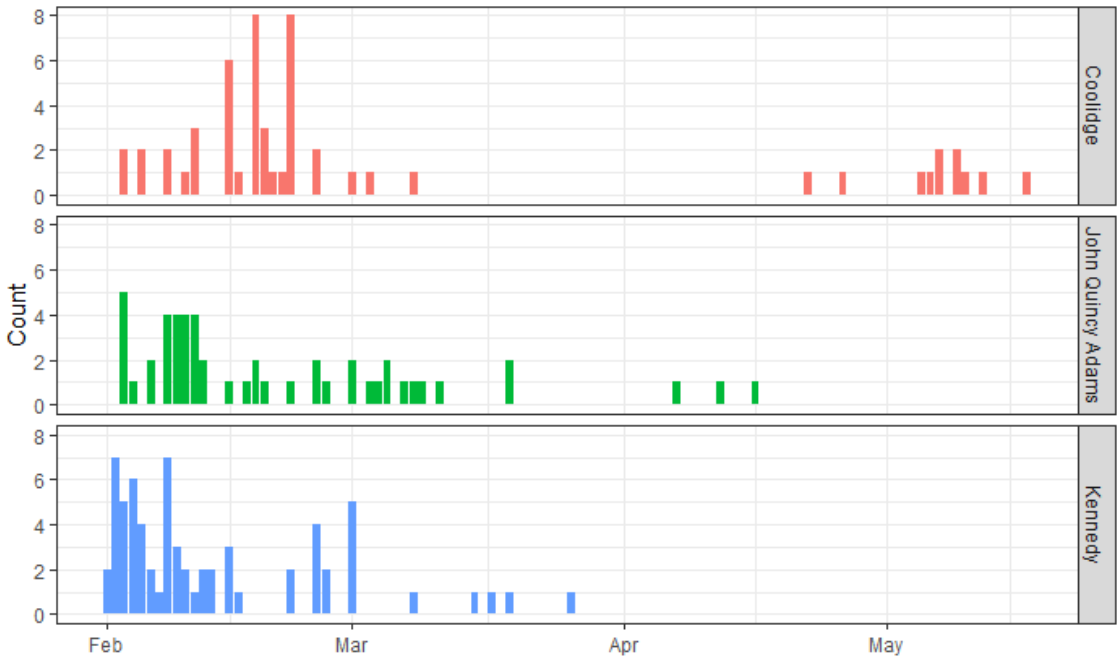

Figure 2.1. COVID-19 case counts by residential dormitory building, 2021

2.2 Space-Time Permutation Scan Statistic

We used the retrospective Space-Time permutation scan statistic contained in the SaTscan program (M. Kulldorff, SaTScan™, v. 10.0.1, October 2021) to retrospectively detect clusters of SARS-CoV-2 infections that appeared during the time period of interest. This method requires no controls or background population data, instead it compares the number of observed cases in a cluster with the expected cases in a cluster, given that there was no space-time interaction (SaTScan guide). In other words, this model detects geographic areas with a higher-than-expected number of cases in a certain time period compared to other geographic areas at that time.

In a two-dimensional space, SaTScan uses a cylinder as the scanning window, with the circular base representing the geography of the cases and the height of the cylinder representing the time period (Xu and Beard 2021). In three-dimensional space, the scanning window is a sphere, with time projected into a fourth dimension. Each scanning window is centered on each case, and the temporal and spatial window

is adjusted and checked until it reaches set boundaries. In the case of our study, the temporal boundaries were a minimum of one day to a maximum of 50% of the study period. The spatial boundaries were restricted to at most 50% of the population-at-risk. Clusters were required to have at least 2 cases to be flagged. The scan statistic was also adjusted for the covariates case gender and day of the week to adjust for testing and reporting artifacts. We conducted the analysis on each dorm building individually, as well as all three buildings in aggregate.

The space-time permutation scan statistic is proposed and further described in M. Kulldorf et al (2005). As with other scan statistics, the space time permutation model detects areas in space and time of unexpectedly high case counts. However, unlike other scan statistics it does not require a uniform population at risk, a control group, or other baseline information. It differs from other scan statistics by its probability model. The expected number of cases is determined using only the cases. In this case C is the number of observed cases and c_{zd} is the number of cases in area z during day d.

$$C = \sum_{z} \sum_{d} c_{zd}$$

For each area and day, we calculated the expected number of cases μ_{zd} conditioning on the observed marginals.

$$\mu_{zd} = \frac{1}{C}(\sum_{z} c_{zd})(\sum_{d} c_{zd})$$

This value is the proportion of all cases in area z times the total cases during day d. μ_A is the expected number of cases inside the scanning window A.

$$\mu_A = \sum_{(z,d)\epsilon A} \mu_{zd}$$

We assume that the probability of a case being in an area z given that it was observed on day d is the same for all days d. When there is no space-time interaction c_A is distributed according to the hypergeometric function with mean μ_A.

$$P(C_A) = \frac{(\sum_{z\epsilon A} c_{zd}) \binom{C - \sum_{z\epsilon A} c_{zd}}{\sum_{d\epsilon A} c_{zd} - c_A}}{\binom{C}{\sum_{d\epsilon A} c_{zd}}}$$

When $\sum_{z\epsilon A} c_{zd}$ and $\sum_{d\epsilon A} c_{zd}$ are small compared to C, c_A is approximately Poisson distributed with mean μ_A. Since c_A is approximately Poisson distributed, we can use the Poisson generalized likelihood ratio (GLR) to measure if the scanning window contains an outbreak.

$$(\frac{c_A}{\mu_A})^{c_A}(\frac{C - c_A}{C - \mu_A})^{C - c_A}$$

Hypothesis testing is done by creating a large number of random permutations of each case in the data set. In our case we generated 99,999 simulated datasets and evaluated statistical significance using Monte Carlo hypothesis testing. The p-value of each cluster is p = R/(S +1) where R is the rank of the maximum GLR from the real data set and S is the number of simulated data sets. So if the GLR from the real data set is higher than the 5000th highest GLM of the simulated data set, it is statistically significant at the 0.05 level.

We conducted three separate analyses for different levels of spatial detail. The most detailed analysis was at the room level, each case was mapped and analyzed by the room the infected student lived in at the time of infection (Figure 2.1). For the second analysis, cases were analyzed at the floor level, meaning that all cases on a floor were grouped to the same location and the location from floor to floor only varied by height (Figure 2.2). Finally, an address-level scan was conducted, where every case in each building was grouped to the same point.

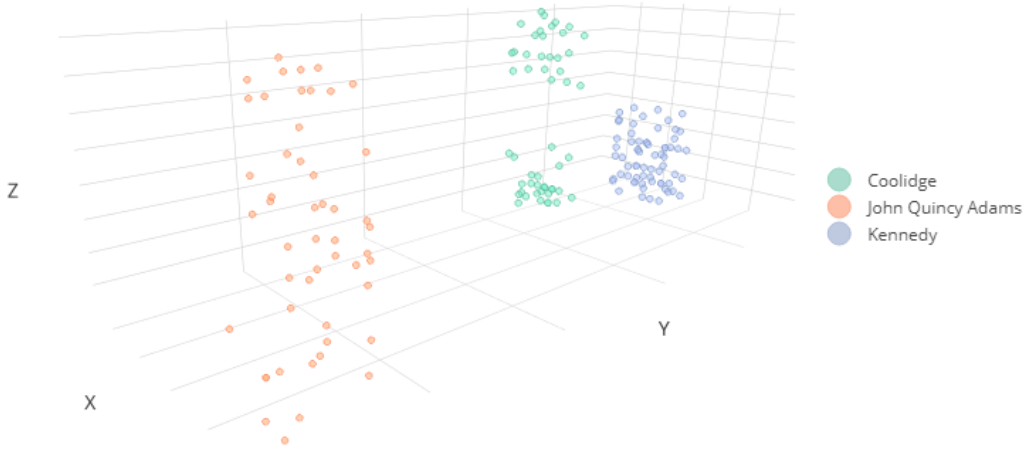

Figure 2.2. Room-Level 3D Plot of SARS-CoV-2 infections in residential dormitories, 2021

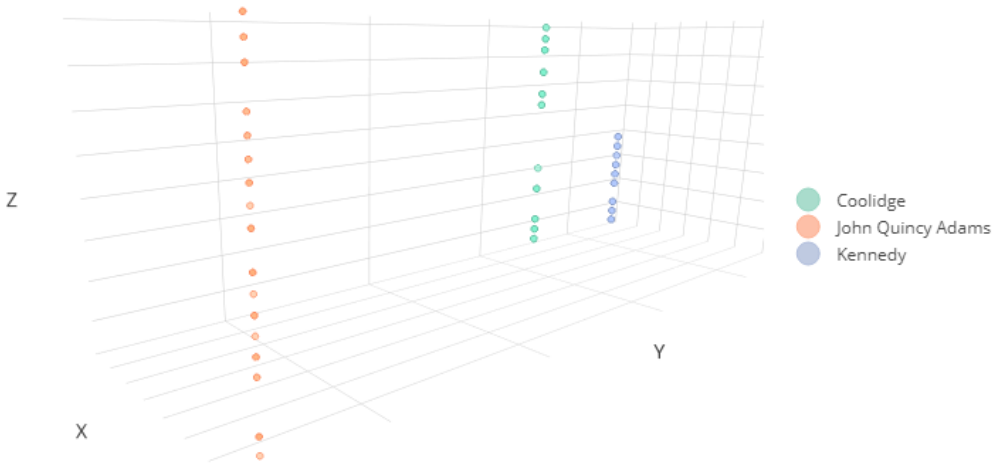

Figure 2.3. Floor-Level 3D Plot of SARS-CoV-2 infections in residential dormitories, 2021

2.3 Purely Temporal analysis with CDC Early Aberration Reporting System (EARS)

As a comparison to the spatial-temporal scan statistic, we used the CDC Early Aberration Reporting System as contained within the "surveillance" package in R (https://cran.r-project.org/web/packages/surveillance/index.html) to detect outbreaks in a purely temporal manner. This statistical aberration detection method was chosen because does not require lengthy historical data to be used as a control and instead uses the previous 5 days to calculated the mean and standard deviation required for the analysis (Hutwagner et all 2005). This method was more appropriate for our analysis because historical data were unavailable due to the recent emergence of COVID-19. While developed for syndromic surveillance (groups of symptoms, as opposed to defined diagnoses) it has been widely used for a range of confirmed infections. We conducted our analysis on the cases of all three dorms in aggregates as well as the cases in all three dorms individually. Using this temporal technique on individual dorms introduces a spatial element to this analysis.

CHAPTER 3

RESULTS

3.1 Room-Level Analysis

Room-Level analysis yielded three significant clusters. Information about these clusters is contained in table 3.1. These clusters were contained within the Coolidge dorm, save for one case in the third cluster which occurred in Kennedy. The cases in each cluster varied from 5 to 8 different floors with lengths of 7 to 25 days.

Table 3.1. Room-level clusters of SARS-CoV-2 infections detected using SaTScan's space time permutation mode, 2021

Cluster	Dorm	Start Date	End Date	Duration (Days)	Radius (Meters)	# of Floors	Observed Cases	Expected Cases	Relative Risk	Test Statistic	p-value
1	Coolidge	20-Apr	15-May	25	33.67	5	10	2.11	4.73	7.841271	0.0009
2	Coolidge	5-May	15-May	10	43.04	8	9	2.14	4.21	6.211861	0.03148
3	Coolidge /Kennedy	15-Feb	22-Feb	7	36.35	6	21	9.15	2.29	6.037012	0.04464

3.2 Floor-Level Analysis

Analysis at the Floor-Level also yielded three significant clusters. All of the clusters were contained within the Coolidge dorm. These clusters had cases from 2 to 7 floors and had a duration of 7 to 25 days.

Table 3.2. Floor-level clusters of SARS-CoV-2 infections detected using SaTScan's space time permutation model

Cluster	Dorm	Start Date	End Date	Duration (Days)	Radius (Meters)	# of Floors	Observed Cases	Expected Cases	Relative Risk	Test Statistic	p-value
1	Coolidge	20-Apr	15-May	25	22.8	5	10	2.16	4.62	7.657971	0.00022
2	Coolidge	15-Feb	22-Feb	7	7.6	2	20	8.37	2.39	6.223845	0.00248
3	Coolidge	5-May	15-May	10	34.2	7	9	2.14	4.21	6.211861	0.00254

3.3 Address-Level Analysis

There were two significant address-level clusters, one in Coolidge and the other in John Quincy Adams. Both clusters had a radius of 0 meters due to all cases within a dorm being geolocated to a single point and being completely contained within each dorm. The clusters varied from 25 to 45 days and contained cases from 11 to 16 floors.

Table 3.3. Address-level clusters of SARS-CoV-2 infections detected using SaTScan's space time permutation model

Cluster	Dorm	Start Date	End Date	Duration (Days)	Radius (Meters)	# of Floors	Observed Cases	Expected Cases	Relative Risk	Test Statistic	p-value
1	Coolidge	20-Apr	15-May	25	0	11	11	3.47	3.17	5.341485	0.00049
2	John Quincy Adams	28-Feb	14-Apr	45	0	16	14	0.41	1.89	2.454119	0.08863

3.4 CDC EARS

For all three dorms, retrospective surveillance using EARS yielded no number of infections above the threshold for an alarm as shown in figure 3.1. Analysis of individual dorms gave an alarm for the Coolidge dorm on February 15th as shown in Figure 3.2, an alarm for the Kennedy dorm on February 25th as shown in figure 3.3, and no alarms for the John Quincy Adams dorm as shown in figure 3.4.

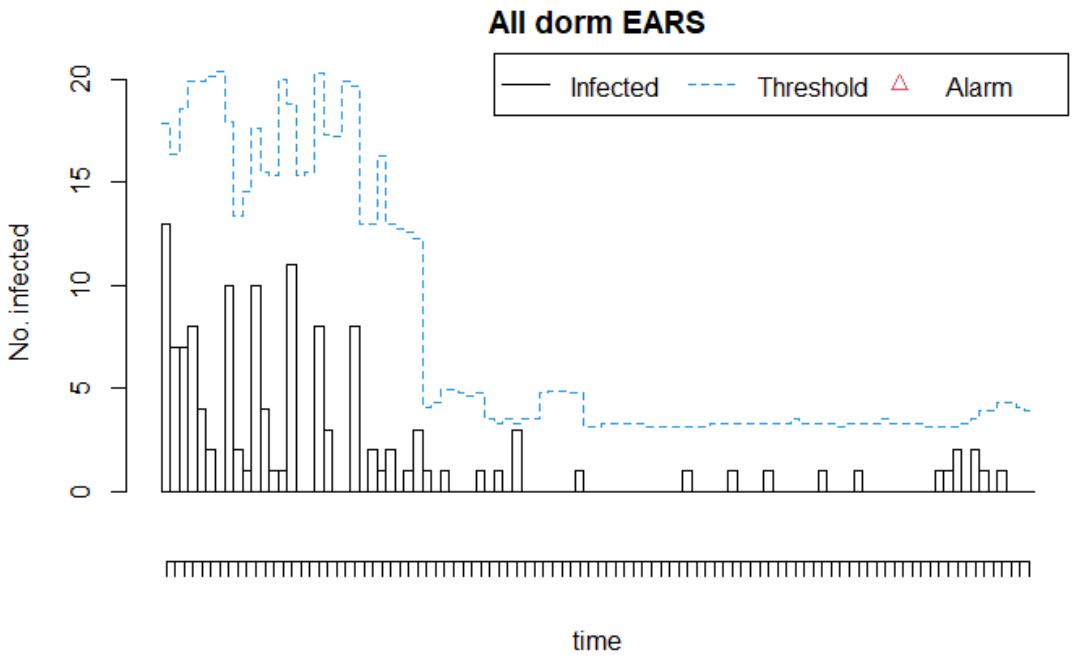

Figure 3.1. CDC EARS analysis of SARS-COV-2 cases in residential dormitories

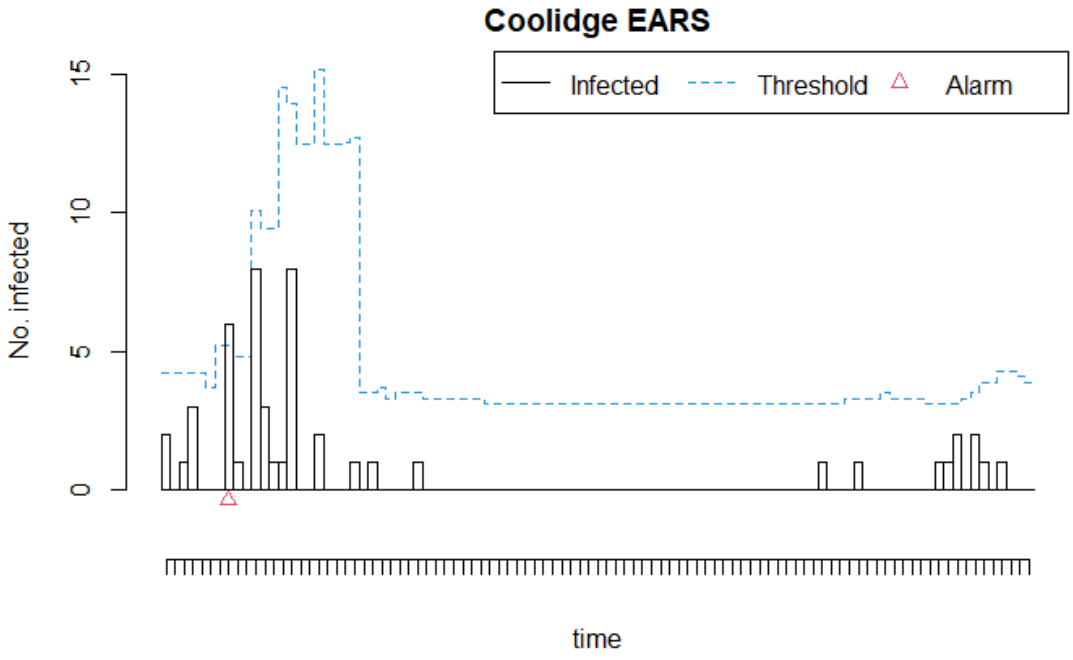

Figure 3.2. CDC EARS analysis of SARS-COV-2 cases in the Coolidge dormitory

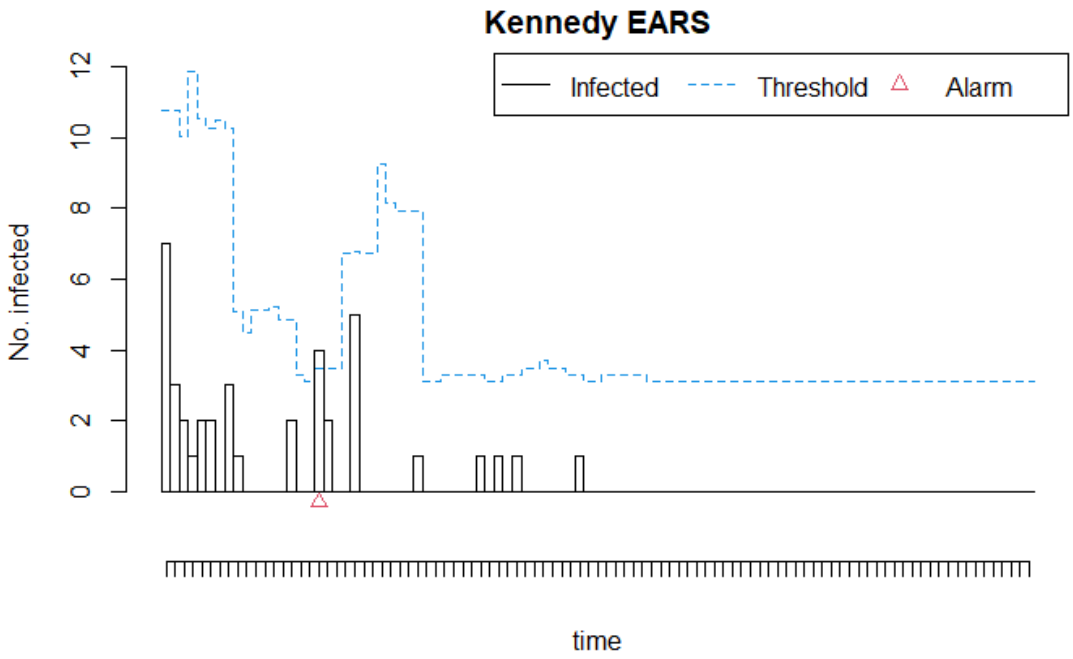

Figure 3.3. CDC EARS analysis of SARS-COV-2 cases in the Kennedy dormitory

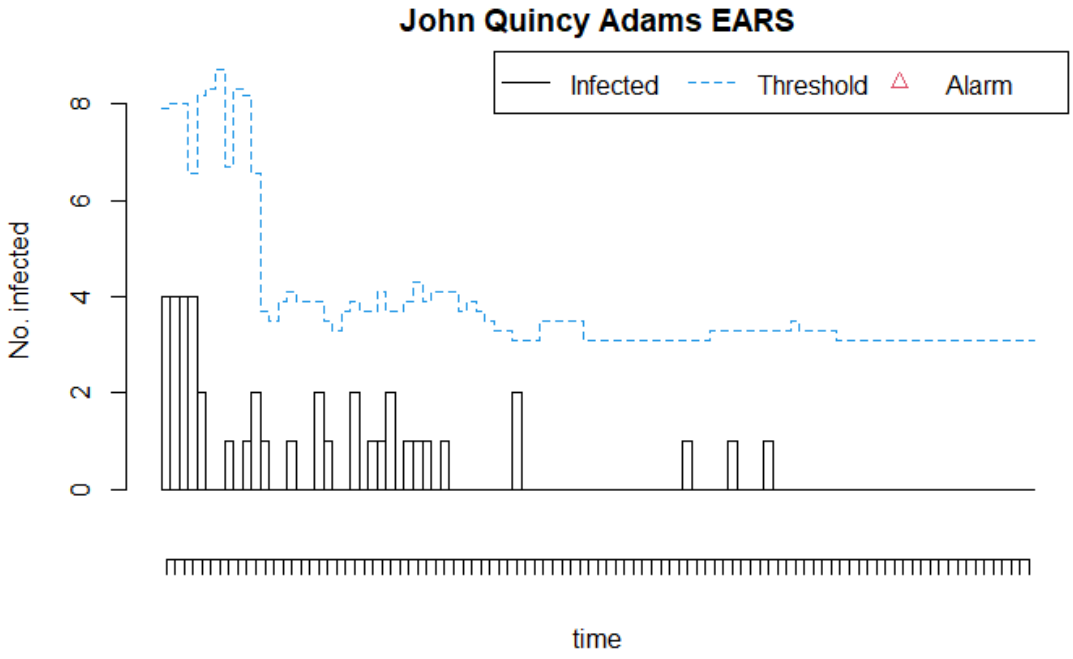

Figure 3.4. CDC EARS analysis of SARS-COV-2 cases in the John Quincy Adams dormitory

CHAPTER 4

DISCUSSION

The goals of retrospective cluster detection are to quantify transmission clusters to gain information to predict future outbreaks, determine possible sources of transmission, and optimize resource allocations for response programs. Implementing spatial-temporal scan statistics at different spatial scales allows us to consider not only these goals, but also to weigh the effectiveness against the complexity, time, and workload involved with analysis at each level of spatial resolution.

4.1 Space-Time Permutation Scan Statistic

Analysis at a dorm room-level provides statistical evidence for three clusters, all of which had cases that were all or mostly within a single dorm (Coolidge). With analysis at an aggregated floor-level, three statistically significant clusters were also detected, two of which were very similar to the clusters detected using the room-level data. However, one cluster was much smaller in size and included cases on only two floors. With analysis at a street address-level, there was evidence for two statistically significant clusters, one in the Coolidge dorm and another in the John Quincy Adams dorm. These address-level clusters are smaller in spatial size than those detected at the other levels of detail, as all the cases in each dorm are at a point centered on each dorm building. The first cluster at the address level was similar to the first clusters detected at the floor and room level, with the same start and end dates, but included cases on more floors, with an estimated lower relative risk. The second cluster at the address level not only occurred in a different dorm than those identified by the other

detail levels, but also had a duration of 45 days, which is much longer than any other significant cluster identified.

Given that each model identified at least some statistically significant clusters within the Coolidge dorm, we can conclude that there were probably multiple localized outbreaks within this building over the study period. While the direct causes of these outbreaks are unclear, the clustering of cases strongly suggests that there was local transmission of infection within the Coolidge building itself. In the future, further steps should be taken to prevent the spread of disease in this potentially high-risk setting.

Direct comparisons between these levels spatial resolution is challenging due to the very localized level at which the analysis occurred, especially when even address-level public health decisions were rarely made during the study period. The number of cases in the Kennedy, Coolidge, and John Quincy Adams dormitories were never large enough to merit a context-specific response. The room-level and floor-level analyses produced comparable results, while the address-level analysis lacked the sensitivity to identify different clusters within the same dorm. With the SaTScan program, analysis for the room-level took 54 minutes to run sufficient iterations. In comparison, the floor-level analysis took just under 5 minutes, and the address-level analysis took 18 seconds. The room-level analysis also required a comparatively large amount of work to geolocate each room and a much longer data processing time than the floor-level analysis. Moreover, students on the same floor are likely to be epidemiologically very similar to each other in this context, as they share bathrooms and communal living spaces. The floor-level analysis shows similar sensitivity with a fraction of the labor and processing, and would appear to be like the most logical choice for this health surveillance in this context.

4.2 CDC EARS

The simpler temporal testing using EARS algorithm detected no statistically significant clustering of cases during the study period when conducted on the cases in all three dorms in aggregate, likely because the case count 5-day running means were monotonic and decreasing during this time period. This demonstrates the importance of the spatial elements of an outbreak. Even appropriately calibrated, a purely temporal analysis can miss relevant epidemiological information. When used on the cases on individual dorms, EARS did give alarms for both the Coolidge and Kennedy dorms. However, this adds a spatial element to the purely temporal EARS and does not adjust for multiple testing. The first date and location flagged by EARS, February 15th in the Coolidge dorm, occurs at the same time as a cluster found by the floor-level and room-level analyses. The other date and location flagged by EARS, February 25th in the Kennedy dorm, does not coincide with a cluster detected in any SaTScan analysis.

4.3 Limitations

Our analysis was limited in many aspects, one of which is scale. The three dormitories studied are near many other dormitories in the Southwest area, and students mixed on many different levels. Some of these other dormitories are closer to the three analyzed dorms than they are to each other. This means that most cross-dorm clustering may be undetected despite being relevant in a public health context. Additionally, the Space-Time Permutation model adjusts for purely spatial and temporal clusters, so clusters are detected comparatively, meaning that they are high numbers of cases compared to other locations in the three studied dorms. This means that if the three analyzed dorms were not representative of the entirety of campus, the results are less generalizable.

Another limitation of the study is the implied spatial relationship between the cases. The analysis was done in a Euclidean manner, represented in three-dimensional space with the absolute difference in height between the floors also being the distance between cases in height. However, viruses do not spread through ceilings and floors (generally), nor are students able to move as easily in the z direction as they are in the x and y direction. To go from one floor to the next takes a much longer time than walking two rooms down the hall, even if the model considers these distances as being identical. The true spatial relationships between floors in this context is not properly represented in Euclidean space.

4.4 Recommendations

For future implementations of the SaTScan, we recommend using SaTScan's Poisson model, which incorporates underlying population data into its analysis. Kulldorff et al report that when high-quality population-at risk data are available, the Poisson-based scan statistic model is expected to perform better than the space-time permutation model. The population-at-risk data is certainly very excellent for the school dormitories, where every student living there is well documented. Additionally, we suggest using a non-Euclidean method of modeling the relationship between floors, similar to the model created by Abboud et al used to study the spread of antibiotic resistant *Klebsiella pneumoniae* in a single clinical site (Abboud et al 2015).

In terms of practical public health surveillance, our findings suggest that the optimal resolution for SaTScan analysis within residence halls is with floor-level data. This level of detail captures case-clustering within buildings but takes far less time geolocating and processing than a room-level analysis. This analysis would complement other early detection strategies such as purely temporal scans, address-level scans, and concurrent wastewater analysis. Building or clusters of buildings with statistically higher case counts or high wastewater viral indicators could be scanned

with SaTScan to identify problematic floors or groups of floors for directed public health response.

BIBLIOGRAPHY

[1] Abboud, C. S., Monteiro, J., França, J. I. D., Pignatari, A. C., Souza, E. E. De, Camargo, E. C. G., Monteiro, A. M. V., Santos, R. G. Dos, and Kiffer, C. R. V. A space–time model for carbapenemase-producing Klebsiella pneumoniae (KPC) cluster quantification in a high-complexity hospital. *Epidemiology & Infection 143*, 12 (Sept. 2015), 2648–2652. Publisher: Cambridge University Press.

[2] Bjorkman, Kristen K., Saldi, Tassa K., Lasda, Erika, Bauer, Leisha Conners, Kovarik, Jennifer, Gonzales, Patrick K., Fink, Morgan R., Tat, Kimngan L., Hager, Cole R., Davis, Jack C., Ozeroff, Christopher D., Brisson, Gloria R., Larremore, Daniel B., Leinwand, Leslie A., McQueen, Matthew B., and Parker, Roy. Higher viral load drives infrequent SARS-CoV-2 transmission between asymptomatic residence hall roommates. Tech. rep., medRxiv, Mar. 2021. Type: article.

[3] Dellicour, Simon, Linard, Catherine, Van Goethem, Nina, Da Re, Daniele, Artois, Jean, Bihin, Jérémie, Schaus, Pierre, Massonnet, François, Van Oyen, Herman, Vanwambeke, Sophie O., Speybroeck, Niko, and Gilbert, Marius. Investigating the drivers of the spatio-temporal heterogeneity in COVID-19 hospital incidence—Belgium as a study case. *International Journal of Health Geographics 20*, 1 (June 2021), 29.

[4] Gorny, Alexander W., Bagdasarian, Natasha, Koh, Azriel Hong Kiat, Lim, Yong Chin, Ong, Jacqueline Soo May, Ng, Bryan Su Wei, Hooi, Benjamin, Tam, Wai Jia, Kagda, Fareed Husain, Chua, Gerald Seng Wee, Yong, Michael, Teoh, Hock Luen, Cook, Alex Richard, Sethi, Sunil, Young, Dan Yock, Loh, Thomas, Lim, Aymeric Yu Tang, Aw, Andrew Kian-Li, Mak, Kenneth Seck Wai, and Fisher, Dale. SARS-CoV-2 in migrant worker dormitories: Geospatial epidemiology supporting outbreak management. *International Journal of Infectious Diseases 103* (Feb. 2021), 389–394.

[5] Greene, Sharon K., Peterson, Eric R., Balan, Dominique, Jones, Lucretia, Culp, Gretchen M., Fine, Annie D., and Kulldorff, Martin. Detecting COVID-19 Clusters at High Spatiotemporal Resolution, New York City, New York, USA, June–July 2020 - Volume 27, Number 5—May 2021 - Emerging Infectious Diseases journal - CDC.

[6] Greene, Sharon K., Peterson, Eric R., Balan, Dominique, Jones, Lucretia, Culp, Gretchen M., Fine, Annie D., and Kulldorff, Martin. Detecting Emerging COVID-19 Community Outbreaks at High Spatiotemporal Resolution — New York City, June–July 2020. Tech. rep., medRxiv, Aug. 2020. ISSN: 2015-6901 Type: article.

[7] Hohl, Alexander, Delmelle, Eric M., Desjardins, Michael R., and Lan, Yu. Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spatial and Spatio-Temporal Epidemiology 34* (Aug. 2020), 100354.

[8] Huang, Chaolin, Wang, Yeming, Li, Xingwang, Ren, Lili, Zhao, Jianping, Hu, Yi, Zhang, Li, Fan, Guohui, Xu, Jiuyang, Gu, Xiaoying, Cheng, Zhenshun, Yu, Ting, Xia, Jiaan, Wei, Yuan, Wu, Wenjuan, Xie, Xuelei, Yin, Wen, Li, Hui, Liu, Min, Xiao, Yan, Gao, Hong, Guo, Li, Xie, Jungang, Wang, Guangfa, Jiang, Rongmeng, Gao, Zhancheng, Jin, Qi, Wang, Jianwei, and Cao, Bin. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet 395*, 10223 (Feb. 2020), 497–506. Publisher: Elsevier.

[9] Hutwagner, Lori, Thompson, William, Seeman, G. Matthew, and Treadwell, Tracee. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). *Journal of Urban Health : Bulletin of the New York Academy of Medicine 80*, Suppl 1 (Mar. 2003), i89–i96.

[10] Kan, Zihan, Kwan, Mei-Po, Wong, Man Sing, Huang, Jianwei, and Liu, Dong. Identifying the space-time patterns of COVID-19 risk and their associations with different built environment features in Hong Kong. *Science of The Total Environment 772* (June 2021), 145379.

[11] Kanankege, Kaushi S. T., Alvarez, Julio, Zhang, Lin, and Perez, Andres M. An Introductory Framework for Choosing Spatiotemporal Analytical Tools in Population-Level Eco-Epidemiological Research. *Frontiers in Veterinary Science 7* (2020).

[12] Kulldorff, Martin, Heffernan, Richard, Hartman, Jessica, Assunção, Renato, and Mostashari, Farzad. A Space–Time Permutation Scan Statistic for Disease Outbreak Detection. *PLOS Medicine 2*, 3 (Feb. 2005), e59. Publisher: Public Library of Science.

[13] Leung, Nancy H. L. Transmissibility and transmission of respiratory viruses. *Nature Reviews Microbiology 19*, 8 (Aug. 2021), 528–545. Number: 8 Publisher: Nature Publishing Group.

[14] Li, Juan, Lai, Shengjie, Gao, George F., and Shi, Weifeng. The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature 600*, 7889 (Dec. 2021), 408–418. Number: 7889 Publisher: Nature Publishing Group.

[15] Rosillo, Nicolás, Del-Águila Mejía, Javier, Rojas-Benedicto, Ayelén, Guerrero-Vadillo, María, Peñuelas, Marina, Mazagatos, Clara, Segú-Tell, Jordi, Ramis, Rebeca, and Gómez-Barroso, Diana. Real time surveillance of COVID-19 space and time clusters during the summer 2020 in Spain. *BMC Public Health 21*, 1 (May 2021), 961.

[16] Shi, Zhicheng, and Pun-Cheng, Lilian S. C. Spatiotemporal Data Clustering: A Survey of Methods. *ISPRS International Journal of Geo-Information 8*, 3 (Mar. 2019), 112. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[17] Xu, Fuyu, and Beard, Kate. A comparison of prospective space-time scan statistics and spatiotemporal event sequence based clustering for COVID-19 surveillance. *PLOS ONE 16*, 6 (June 2021), e0252990. Publisher: Public Library of Science.