A Web-based Spatial Decision Support System of Wastewater Surveillance for COVID-19 Monitoring: A Case Study of a University Campus

Wenwu Tang[a,b*], Tianyang Chen[a,b], Zachery Slocum[a,b], Yu Lan[a,b], Eric Delmelle[a,b,c], Don Chen[d], Neha Mittal[e], Jacelyn Rice-Boayue[d], Tarini Shukla[b,f], Sophia Lin[f,g], Srinivas Akella[h], Jessica Schlueter[e], Mariya Munir[g], Cynthia Gibas[e]

[a] Department of Geography and Earth Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223 USA

[b] Center for Applied Geographic Information Science, University of North Carolina at Charlotte, Charlotte, NC 28223 USA

[c] Department of Geographical and Historical Studies, University of Eastern Finland, Joensuu, Finland

[d] Department of Engineering Technology and Construction Management, University of North Carolina at Charlotte, Charlotte, NC 28223 USA

[e] Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC 28223 USA

[f] Infrastructure and Environmental System Program, University of North Carolina at Charlotte, Charlotte, NC 28223 USA

[g] Department of Civil & Environmental Engineering, University of North Carolina at Charlotte, Charlotte, NC 28223 USA

[h] Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC 28223 USA

*Corresponding author, Email: WenwuTang@uncc.edu, Phone: 1-704-687-5988

Abstract

The ongoing COVID-19 pandemic has produced substantial impacts on our society. Wastewater surveillance has increasingly been introduced to support the monitoring, and thus mitigation, of COVID-19 outbreaks and transmission. Monitoring of buildings and sub-sewershed areas via a wastewater surveillance approach has been a cost-effective strategy for mass testing of residents in congregate living situations such as universities. A series of spatial and spatiotemporal data are involved with wastewater surveillance, and these data must be interpreted and integrated with other information to better serve as guidance on response to a positive wastewater signal. The management and analysis of these data poses a significant challenge, in particular, for the need of supporting timely decision making. In this study, we present a web-based spatial decision support system framework to address this challenge. Our study area is the main campus of the University of North Carolina at Charlotte. We develop a spatiotemporal data model that facilitates the management of space-time data related to wastewater surveillance. We use spatiotemporal analysis and modeling to discover spatio-temporal patterns of COVID-19 virus abundance at wastewater collection sites that may not be readily apparent in wastewater data as they are routinely collected. Web-based GIS dashboards are implemented to support the automatic update and sharing of wastewater testing results. Our web-based SDSS framework enables the efficient and automated management, analytics, and sharing of spatiotemporal data of wastewater testing results for our study area. This framework provides substantial support for informing critical decisions or guidelines for the prevention of COVID-19 outbreak and the mitigation of virus transmission on campus.

Keywords: Wastewater surveillance, spatial decision support systems, COVID-19, Web GIS

1. Introduction

The COVID-19 pandemic has fueled a renewed interest in wastewater-based epidemiology. Wastewater testing for traces of viral and bacterial pathogens has been used for decades to track and manage outbreaks of infectious disease (Prado et al., 2012; Tambini et al., 1993). Early reports in mid-2020 demonstrated that wastewater concentrations of SARS-CoV-2 could serve as a leading indicator for cases detected by clinical testing within city sewersheds (Ahmed et al., 2021; Peccia et al., 2020), with collection of samples from wastewater treatment plant influent providing coverage of entire cities or large neighborhoods. The practical application of monitoring at city scale is primarily to detect infection trends in communities, which has been especially useful in the case of COVID-19, both because COVID-19 infections may be asymptomatic for several days prior to detection of cases by testing, and because especially in the early months of the pandemic, testing capacity lagged behind the rapid spread of the disease. In such scenarios, wastewater testing can serve as a leading indicator of the increase of disease incidence in an urban area. There has also been an increasing interest in monitoring in neighborhood or smaller scale areas for the presence of the SARS-CoV-2 virus in wastewater, because such small-scale monitoring can provide evidence to support targeted public health interventions including distribution of masks or selection of populations for increased testing (Bowes et al., 2021).

COVID-19 is easily transmitted in congregate living situations, with early and devastating outbreaks being documented in nursing homes and jails (Kırbıyık et al., 2020; Lam-Hine et al., 2021). Beside these, other indoor settings such as schools (including universities), restaurants, and hospitals have been identified as having high risk for the spread of COVID-19 (Fox et al.,

2021; Lam-Hine et al., 2021). Many universities attempted to implement some type of wastewater surveillance program during the early months of the pandemic, with varying degrees of success (Gibas et al., 2021; Harris-Lovett et al., 2021; Karthikeyan et al., 2021). To effectively detect and monitor outbreaks of COVID-19 in these indoor settings requires wastewater surveillance capabilities at small spatial scales such as building level. The study reported in this article is focused on building-level wastewater surveillance for COVID-19 monitoring from a spatiotemporally explicit perspective.

Wastewater surveillance typically requires a set of sequential steps, including sample site setup, sample collection (including storage and shipping; per CDC Wastewater Surveillance strategy), lab analysis, and subsequent analysis and visualization of wastewater testing results and associated data. Geographic Information Systems (GIS) methods have been applied for the management and mapping of spatially explicit data related to wastewater testing and COVID-19 monitoring, and dashboard techniques have gained increasing attention due to their visual presentation capabilities within web-based environments (Dong et al., 2020; Lan et al., 2021; Naughton et al., 2021). Yet, most of the existing dashboards for COVID-19 and wastewater studies only concentrate on management and visualization of relevant spatial or spatiotemporal data; their support on spatial analytics and modeling capabilities is inadequate. Spatial analytics and modeling, however, are pivotal in discovering patterns of interest hidden in complicated spatiotemporal data, and providing predictive or scenario analysis capabilities for monitoring and mitigation of pandemic situations (Franch-Pardo et al., 2020). Spatial Decision Support Systems (SDSS) hold potential in filling this gap.

SDSS, which originated from the domain of Geographic Information Science (Armstrong et al., 1986; Sugumaran & Degroote, 2010), have been increasingly applied to assist with decision making within spatially explicit contexts. SDSS is based on (but more than) the integration of decision support systems and GIS, and provides inherent support for spatial analytics and modeling capabilities. This makes SDSS unique and powerful in informing decision making processes associated with complex spatial or spatiotemporal phenomena. A variety of applications such as environmental monitoring, natural resources, public health, transportation, and land use and land cover change have built SDSS to address complex decision problems within spatially explicit contexts (Delmelle et al., 2014; Keenan & Jankowski, 2019; Sugumaran & Degroote, 2010). In particular, driven heavily by Internet technologies and cyberinfrastructure (NSF, 2007), web-based SDSS has received much attention over the past few years (Lan et al., 2020; Lee et al., 2017; Tayyebi et al., 2016). While a growing body of the literature has highlighted the power of web-based SDSS, the applications of web-based SDSS for the resolution of complex spatiotemporal decision problems in general and small-scale wastewater surveillance for COVID-19 monitoring, in particular, remain scant.

In this article, we describe a web-based SDSS framework for building-level wastewater surveillance. We used a university campus (the main campus of the University of North Carolina at Charlotte) as a study case. This framework supports the automated synchronization and update of lab test results, space-time cluster analysis for identifying hotspots of COVID-19 incidents at the building level over time, and automated update of dashboards within web-based environments. The integration of these geospatial data and analytics capabilities play a critical role in providing timely information on COVID-19 incidents in the study region over time.

Specifically, we focus on addressing the following sets of research questions in this study: 1) Are there any space-time clusters of positive wastewater testing results at the building level and where are they? 2) What are those sampling sites that exhibit similar responses over time in terms of wastewater testing results and where are they?

The remainder of this article is organized using the following structure. In section 2, we discuss the background and relevant literature of this study. In section 3, we present the study area and data, the design of the entire web-based SDSS framework as well as its implementation. Section 4 presents the results including space-time cluster analysis, and Section 5 gives relevant discussion. Section 6 concludes this article.

2. Literature Review

2.1. Wastewater Surveillance

A typical workflow for building-level wastewater surveillance includes collection of a sample at regular intervals with laboratory results within 24 hours of collection. Samples can be collected using a variety of methods (Medema et al., 2020), ranging from collection of a sample volume at one timepoint (a “grab” sample), to composites collected by passive sampling for example using fibrous swabs (Liu et al., 2021), and composites collected using pump autosamplers which add to the sample at regular intervals over the course of a day prior to collection. Once collected, samples are processed and concentrated. A wide variety of methods are available for this concentration step as well, and choice of method is governed by a combination of viral recovery efficacy, cost, materials availability, and processing time, as described in our previous work (Juel et al., 2021). RNA is extracted from the concentrated sample, and virus is quantified using

a molecular detection protocol such as RT-qPCR or RT-ddPCR (Barua et al., 2021; Ciesielski et al., 2021), which provides a viral concentration in terms of copies of virus per liter of wastewater collected. This value can be used effectively as a simple binary indicator of positivity, as demonstrated in the pilot phase of our campus monitoring program (Gibas et al., 2021) but also has the potential to connect the information about population size and volume of water used in the building to provide an estimate of the number of individuals who might be SARS-CoV-2 positive (Sweetapple et al., 2022). Once a positive signal is detected, a decision is made about whether to test all individuals in that building, after consulting institutional information about individuals who have recently tested positive or been connected to that site via contact tracing. If there are no previously-known individuals who are likely to be the source of the positive signal, then the entire building population is subjected to clinical testing.

While many institutions and localities have deployed wastewater testing for SARS-CoV-2 during the pandemic, only a small fraction of these projects have so far made data available in service of larger efforts to develop quantitative models and consistent practices in wastewater epidemiology (Naughton et al., 2021). Data dashboards are a common means for sharing such information when it is made available, and in some cases have been incorporated into state-level public health reporting (e.g., see https://covid19.ncdhhs.gov/dashboard/wastewater-monitoring). Dashboard techniques have been extensively applied for the sharing of data related to COVID-19. A number of dashboards have been developed and deployed to support the wastewater surveillance initiatives for the monitoring of COVID-19 worldwide. For example, there are a number of dashboards registered via the web site of COVIDPoops19 project (Naughton et al., 2021). About 40% of these dashboards have built-in Web GIS functionality. The software

platforms used to present these dashboards include Esri ArcGIS Online, Tableau, R Shiny, Microsoft Power BI, and CARTO. The first three (ArcGIS Online, Tableau, and R Shiny) are the dominant choices for the implementation of wastewater surveillance dashboards. Most of the wastewater data managed and reported by these wastewater dashboards are at the wastewater treatment plant level and collected weekly, while a smaller number of projects report daily or multiple days per week. A few universities make campus wastewater data available in real time via public dashboards (e.g. University of California at San Diego, Clemson University), but in other cases, for instance at the University of North Carolina at Charlotte, the concern of upper administration not to alarm students or parents with details of wastewater alerts has resulted in a decision to keep this information for internal use only. A number of existing dashboards only focus on the visual presentation (in maps or charts) of wastewater-related data, and may not provide the spatiotemporal analytics and modeling of wastewater testing results and relevant data. The need for spatiotemporal analysis and modeling to guide the study of wastewater testing results for the monitoring of COVID-19 outbreak and prevention has been recognized in the literature (Karthikeyan et al., 2021).

2.2. Spatial Decision Support Systems

SDSS are integrative computer-based systems that provide decision-making support for complex spatial problems via the fusion of spatial data management, modeling, and visualization capabilities (Densham, 1991; Malczewski, 1999; Sugumaran & Degroote, 2010). SDSS, with an origin from Decision Support Systems (Marakas, 2003), are distinguished by their ability to handle decision-making support within a spatially explicit context via the incorporation of GIS-based functionality. Yet, SDSS differ from GIS in that they encompass spatial modeling

capabilities to aid decision-making (Armstrong et al., 1986; Sugumaran & Degroote, 2010). For example, with the incorporation of a spatial simulation model, SDSS can enable what-if scenario analysis to explore potential alternative solutions of a spatial problem. The spatial optimization model helps SDSS identify spatially explicit optimal solutions facing decision makers (represented by site selection problems). Further, spatial statistical models allow for the discovery of spatial patterns of interest (e.g., clusters of disease or accidents) from spatial data. All these modeling capabilities can be built within a SDSS that informs and facilitates decision making processes associated with complex spatial or spatiotemporal problems (Ghosh, 2008). In terms of implementation, a SDSS includes the following functional modules: data management, model management, visualization and report generation, and a user interface (Armstrong et al., 1986; Densham, 1991; Sugumaran & Degroote, 2010).

While the study of SDSS in early stages focuses on the development of conceptual architecture, cyberinfrastructure-enabled computing technologies such as web and cloud computing have been fostering the implementation and applications of SDSS into different domain studies (Sugumaran & Degroote, 2010; Tang et al., 2017). For example, Mwaura and Kada (2017) presented a web-based SDSS in which a multi-criteria decision making model was used to evaluate potential sites of geothermal wells in Kenya, east Africa. Crimi et al. (2019) investigated the identification of priority regions in Bradford, UK for freight lorry parking within a web-based SDSS environment. Lan et al. (2020) applied web-based SDSS that guides the monitoring and sharing of water quality information of private wells in Gaston County, NC, USA. Spatial interpolation algorithms were used in Lan et al.'s work to generate the spatially continuous distribution of water quality that will inform residents or governments for potential water contamination.

3. Materials and Methods

3.1. Study Area and Data

Our study area is the main campus of the University of North Carolina at Charlotte, USA (see Fig. 1). The main campus of the University (35°18'25"N, 80°44'06"W) is located in the north of the City of Charlotte (within Mecklenburg County). The University is an urban university with about 3,000 employees (including faculty and staff); and 30,146 students in the Fall semester of 2020. Among them, around 6,000 students are living in residential halls on campus. In total, there are 138 buildings in the main campus, 33 residence halls, 32 academic buildings, and 73 other types. Please see Appendix 1 for sources of the aforementioned information about the University. In terms of topography, the main campus is high in east and west and low in the middle (range of elevation: 176-226 meters). The slope of the main campus varies from 0^{o} to 25^{o} (based on a 1-m DEM derived from LiDAR point cloud data). The Toby Creek area is the lowest-lying region on campus. Toby Creek flows through the campus and discharges into Mallard Creek at the north end of the campus. The university's sewer system is composed of gravity sewer lines, where a sampling at a specific sewer manhole location will be affected by upstream nodes. Lateral and branch sewer lines collect wastewater from all residence and academic buildings, and then connect to a main sewer line (Charlotte Water's wastewater system) which parallels Toby Creek. Campus wastewater is treated at the nearby Mallard Creek Treatment Facility.

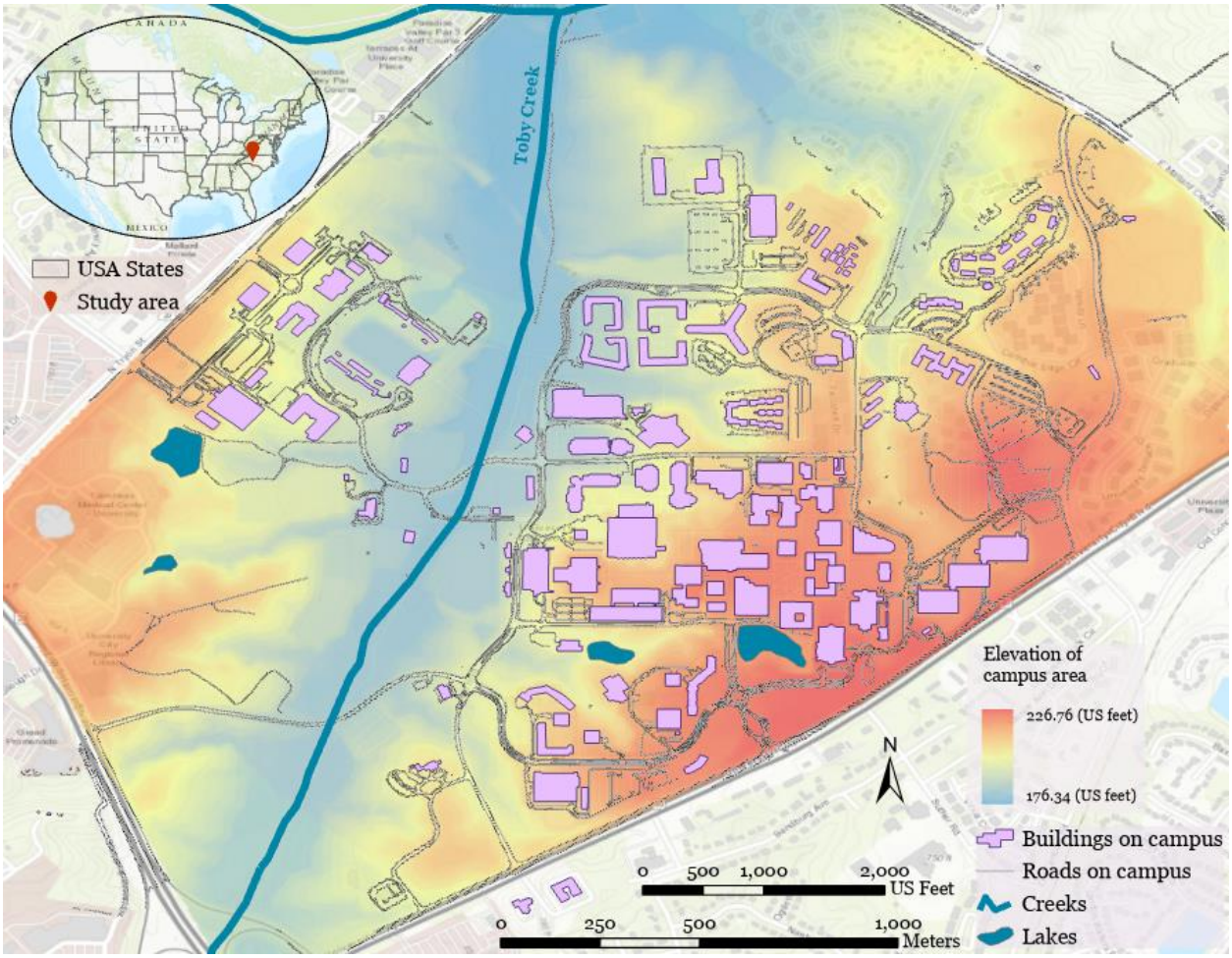

Fig. 1. Map of the main campus of the University of North Carolina at Charlotte, USA (sewage network details are not shown for the protection of physical security of university infrastructure).

The University of North Carolina at Charlotte launched its wastewater-based epidemiology (WBE) surveillance program in late Summer 2020 to assist the University in monitoring COVID-19 incidence. Wastewater signal has been used since that time to identify dormitory populations for testing ("surge testing") in the event of detection of SARS-CoV-2 virus in the absence of a previously identified source. The wastewater surveillance program has been collecting and analyzing wastewater samples since September, 2020. A team of faculty, staff,

and students from bioinformatics, engineering, computer science, and geography collaborate to develop this monitoring system, with infrastructure support from the University's Facilities Management staff. The WBE team has also developed a Building Information Modeling (BIM; see Becerik-Gerber et al. (2012)) 3D model for each residence hall on campus. Each BIM model includes the building envelope and plumbing fixtures, which can be used to identify rooms and zones in which potential infected individuals are located. Wastewater data collected together with BIM models have allowed campus administration to make timely and targeted decisions to prevent the cluster outbreak and spread of COVID-19 on campus (see Gibas et al., 2021 for detail). We collected spatial data to support the wastewater surveillance work for our study area. These data include buildings, sewer lines, sampling sites, road network, and elevation.

Table 1. Spatial data collected for the wastewater surveillance work for the study area.

Spatial Data	Data source	GIS Data Format
Buildings	Department of Facilities Management of UNC Charlotte	Polygon Vector
Sewer lines	Department of Facilities Management of UNC Charlotte	Polyline Vector
Sampling sites	Wastewater Surveillance Task Force Group at UNC Charlotte	Point Vector
Road network	Department of Facilities Management of UNC Charlotte	Polyline Vector
Elevation	U.S. Geological Survey, 3D Elevation Program	Raster

There are in total 38 sampling sites that were identified and established for wastewater collection since Fall 2020 (see Fig. 2 for illustration). These sampling sites are organized in two types: for residence halls (a sampling site covers a building or part of the building) and for buildings within a sub-sewershed—referred to as neighborhood site in this study (a sampling site covers multiple buildings). Manholes and plumbing cleanouts are selected to set up these sampling sites. As a manhole may connect to multiple sewage lines from different buildings, a manhole may have

multiple auto-samplers with probes deployed in different directions (up to two in our study) installed to collect sewage samples from different buildings. Further, a building (typically large) may have two or more sampling sites each covering different parts of the building. These sampling sites cover in total 89 buildings on campus for wastewater monitoring. We used a Trimble GPS handheld unit (with a submeter accuracy) to obtain the coordinates of the sampling sites. However, 10 of 38 samplers are located either very close to the building or inside the building, which degrades the signal quality of GPS satellites. Therefore, their locations are determined using Google Earth and images taken using a digital camera. One sampling site is completely under trees with dense canopy, where we cannot determine its exact coordinates using a GPS instrument or Google Earth imagery. In such a case, we used the location of the corresponding manhole (identified from the GIS data of the sewage network) as the coordinates of the sampling site.

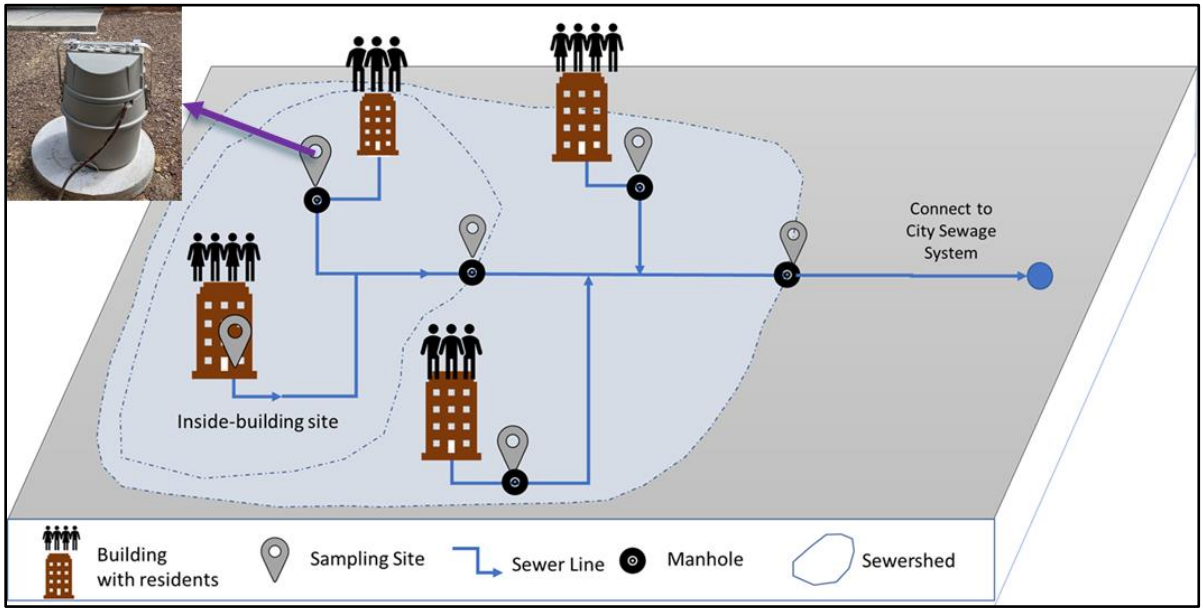

Fig. 2. Illustration of sampling site setup for building-level wastewater surveillance.

3.2. Methods

In this section, we present the framework of the web-based SDSS and its main components. Fig. 3 illustrates the design of the web-based SDSS framework for wastewater surveillance in this study. This framework supports the data management, model management, and visualization of wastewater data that are spatiotemporally explicit. The integration of these functionality allows for the automated synchronization of wastewater testing results, on-demand spatiotemporal analysis of COVID-19 incidents from wastewater results, and automatic update of Web GIS dashboard that supports timely decision making in a spatially explicit manner.

Building-level wastewater surveillance typically includes three steps (see Gibas et al., 2021): collection of wastewater samples, sample concentration and RNA extraction, and detection of COVID-19 virus. Various sample-related data are generated from these steps. These data are characterized with space-time stamps and associated with different sampling sites, buildings, and sewersheds. Fundamentally, these data are space-time series that represent various information related to wastewater testing over space and time. Mathematically, our wastewater surveillance data (noted as W) can be formulated as in Eq. 1:

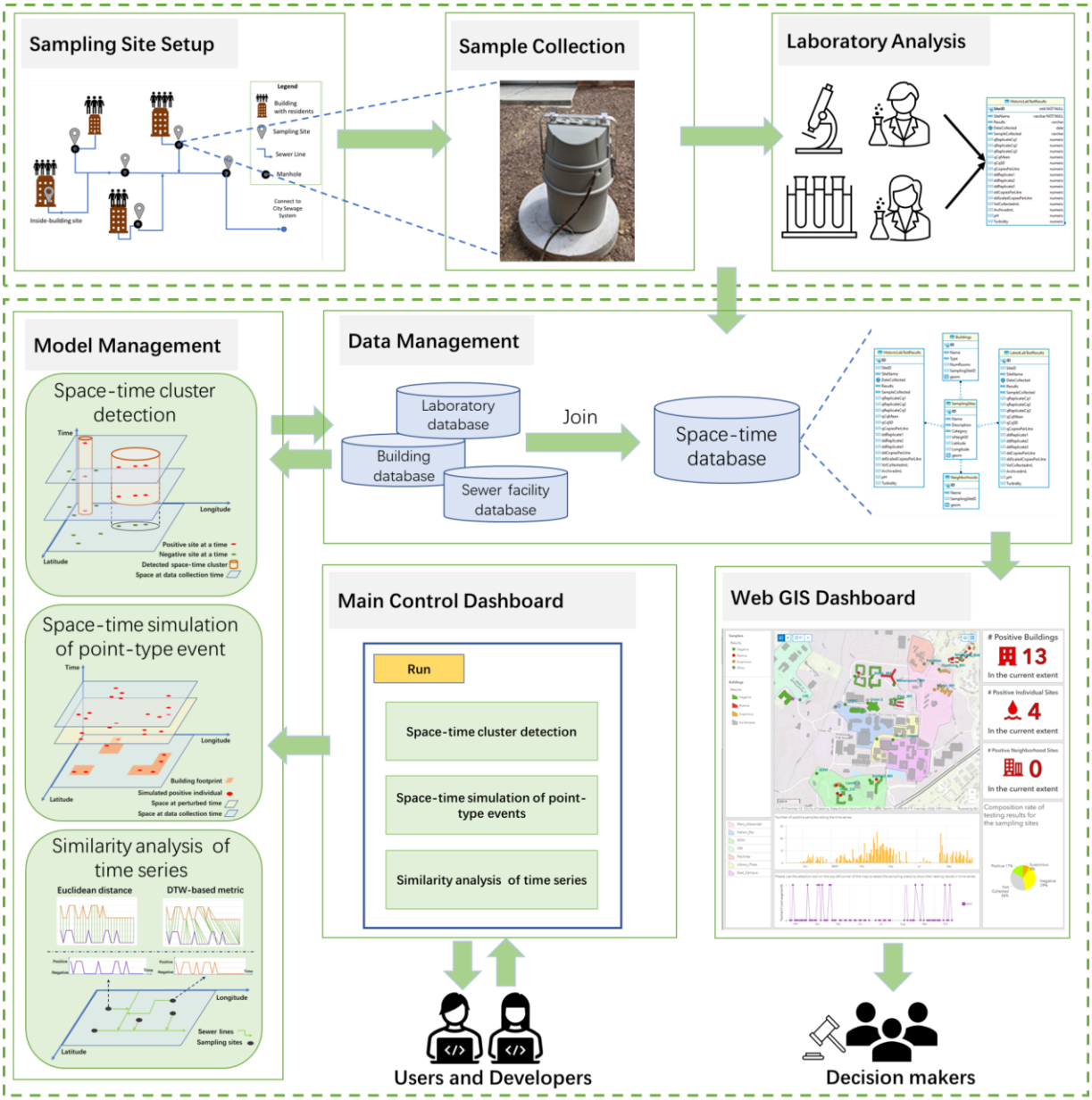

Fig. 3. Framework of the web-based spatial decision support system for wastewater surveillance.

$$W=\{W(i,t) \mid W(i,t) = \{id, w, v_1, v_2, \ldots, v_p\}\} \quad (1)$$

where:

i: sampling site ID, $i \in [1,2,\ldots,n]$;n: number of sampling sites;

t: ID of time step; $t \in [t_1, t_2, \ldots, t_m]$; t_1: beginning date of wastewater sampling; t_m: end date of sampling; m: number of sampling dates;

285 *id*: ID of the sample at site *i* and time *t*.

286 $w(i,t)$: wastewater testing result for site *i* at time *t* ($w(i,t)=\{0,1\}=\{$negative, positive$\}$);

287 $v_1(i,t)$, $v_2(i,t)$, ..., $v_p(i,t)$: all other variables associated with site *i* at time *t;* These variables
288 may change over time or not (e.g., testing results will change over time but the ID of
289 associated building(s) will not).

290 *p*: number of other variables for a sampling site;

291 Among these variables, the wastewater testing result $w(i,t)$ is a binary variable that indicates
292 whether COVID-19 virus is detected (1: positive; 0: negative) for a sampling site on a specific
293 date. In this study, qPCR detection results from three sample replicates are used to determine
294 whether a sample is considered positive or not. When the virus concentration (mean Cq) values
295 of all three sample replicates are lower (indicating higher viral load) than the empirically
296 determined limit of detection threshold, the corresponding wastewater sample is considered
297 positive. For the purposes of determining administrative response on campus, samples must have
298 all three replicates producing signals to be considered "positive". Any samples that have only ⅔
299 replicates producing signals are considered "suspicious" and 1 or fewer replicates producing
300 signals considered negative. This "suspicious" designation is only used for administrative
301 decision purposes. For more detail, please refer to Gibas et al. (2021). In our study here, samples
302 that have 2 or less replicates producing signals are treated as negative (i.e., suspicious and
303 negative samples are merged into a single category: negative).

304 **3.2.1. Spatiotemporal data management and data synchronization**

305 We developed an object-based spatiotemporal data model (see Fig. 4A) to represent
306 spatiotemporally explicit information related to building-level wastewater surveillance for
307 COVID-19 monitoring. Spatiotemporal data models have been developed to represent dynamic
308 geospatial phenomena (Chen et al., 2016; Pelekis et al., 2004; Peuquet & Duan, 1995). Based on

spatiotemporal data models, data structures and databases can be designed and implemented to handle data with spatiotemporal stamps. A series of spatiotemporal data models have been proposed in the literature, including snapshot-based, event-based, and object-based (Pelekis et al., 2004). Our spatiotemporal data model is object-based, in which a spatiotemporal object represents a geospatial entity in space and time. As the geometry of sampling sites and buildings does not change, our spatiotemporal data model only needs to take into account change in attributes (non-spatial information) associated with sampling sites or buildings. Thus, a wastewater sample collected at a site at a specific date is abstracted as a spatiotemporal object associated with a set of variables, including sampling site information (geometry: point), building information (geometry: footprint polygon), and lab testing results. Fig. 4B is the entity-relationship (ER) diagram that we used to build the geodatabase based on the spatiotemporal data model. Database tables were created to manage the spatiotemporal data associated with wastewater surveillance (including sampling sites, buildings, sewersheds, historic lab testing results, and latest lab testing results). Further, we used a set of database tables to maintain the relationships between sampling sites and buildings, as well as sampling sites and sewersheds.

We developed an automated synchronization module to upload wastewater testing results once they are available (including real-time and historic data). This automated data synchronization module is implemented within a web-based interface. This synchronization module takes sample testing results (in a delimited file; CSV format) as input and associates these testing results with corresponding buildings or sewersheds (through SQL style left-joins). Then, these testing results are updated to the spatiotemporal database.

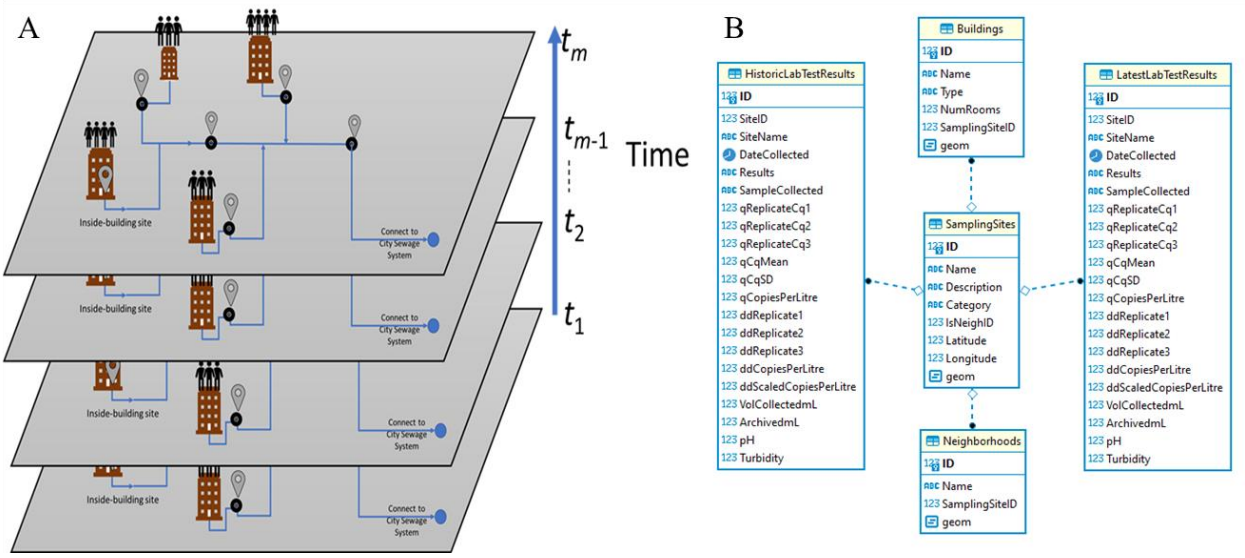

Fig. 4. Illustration of spatiotemporal data model (A) and entity-relationship diagram (B) for building-level wastewater surveillance.

3.2.2. Spatiotemporal analysis of wastewater testing results

To address the research questions aforementioned in the Introduction section requires the use of space-time analysis and modeling approaches. We chose to use space-time scan statistics, space-time simulation of asymptomatic individuals, and similarity analysis of space-time series.

3.2.2.1. Space-time scan for cluster detection

In this study, we utilized space-time scan statistics for the detection of space-time clusters of positive wastewater samples reported from wastewater surveillance. We used Kulldorff's retrospective space-time scan statistic (Kulldorff, 1999; Kulldorff et al., 1998), implemented in SaTScan (version 9.6). A variety of studies have applied the space-time scan statistics approach to detect clusters of covid cases during the COVID-19 pandemic (see, e.g., Desjardins et al., 2020; Hohl et al., 2020; Kim & Castro, 2020; Masrur et al., 2020). However, the space-time

cluster detection for COVID-19 monitoring is often applied at large spatial or jurisdictional scales (e.g., state or county level for a country). To our knowledge, this is the first time that the space-time scan statistic is used to detect the presence of COVID19 in wastewater and at a small spatial scale (building level).

The space-time scan statistics uses a cylinder-based scanning window to detect the cluster of space-time objects (e.g., positive wastewater samples here; see Fig.5). The base of the cylinder defines the geographic region covered by the scanning window (the radius of the base is the spatial bandwidth) while the height represents the time duration of the scanning window (i.e., temporal bandwidth). When applying space-time scan statistics, the center of the cylinder is placed at each spatial object (point-types; centroids can be used for polygon-type objects) and the spatiotemporal bandwidth is varied. Then, by using a likelihood ratio test, the number of observed events within and outside the cylinder is compared against their expected values based on Poisson or Bernoulli models (Kulldorff, 1997). Events within a cylinder scanning window with highest likelihood ratio (indication of elevated risk) are identified as a space-time cluster. Monte Carlo approach can be used to test the significance of the cluster(s). As the wastewater testing results are a binary variable (positive or negative) in this study, we used the Bernoulli model for the probability model used by the space-time scan statistics.

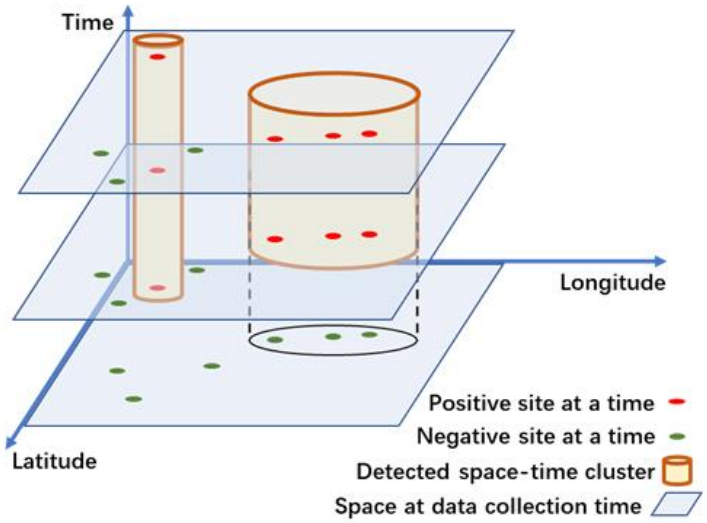

Fig. 5. Illustration of using cylindrical scanning windows for space-time scan statistics.

3.2.2.2. Space-time simulation of asymptomatic individuals

In this study, wastewater testing results from a sampling site are indicative of the situation of the associated building(s)--whether there are presymptomatic individuals in the building. However, the location of the individual(s) within the building is unknown (for privacy protection)—i.e., spatial uncertainty. Further, collected samples on a particular day may be reflective of a prior contamination, keeping in mind that samples were collected every two days or more instead of every day in our study–i.e., temporal uncertainty. Therefore, we used a space-time point pattern simulation approach (see Diggle, 2013) to generate the locations of presymptomatic individuals within the associated building (footprint in polygonal form) and the time that the individuals begin to shed virus. In other words, this approach allows us to simulate space-time locations (where and when) of the presymptomatic individuals, represented as space-time objects in this study.

Fig. 6 illustrates the algorithm of the simulation of space-time point patterns of asymptomatic individuals within buildings. The space-time point pattern simulation begins with footprint polygons of all sampled buildings to determine the spatial location of an asymptomatic individual. A point is randomly generated within the bounding box of the footprint of each building. The point is retained if it is located within the building footprint polygon. Once the spatial location of the presymptomatic individual is determined, the date that the individual begins to shed virus is obtained by randomly perturbing the original sampling date up to *n_perturb* days before (e.g., *n_perturb*=3 in this study). This procedure is applied to each building for a number of Monte Carlo repetitions (e.g., 1,000 repetitions used in this study). After the space-time location is determined, associated sampler site data and testing results are joined. The number of days for perturbation is based on the sampling frequency within a week. For example, 3 days could be used to cover the tri-weekly testing interval. Once simulated results are generated, space-time cluster analysis can be performed on these simulated spatiotemporal point patterns to examine the robustness of space-time clusters detected from observed data.

Algorithm for simulation of spatiotemporal point patterns of asymptomatic individuals

```
Parameters:
   n_monte:   number of Monte Carlo runs
   n_perturb: number of days for temporal perturbation
Begin Algorithm
   For each Monte Carlo run of n_monte repetitions
         For each sampling record (associated with a building and time)
                Randomly generate a point within the building footprint for the sample site;
                Randomly generate the time by perturbing sampling date up to n_perturb days before;
         End for sampling record
   End for Monte Carlo run
End Algorithm
```

Fig. 6. Algorithm of simulation of space-time locations of asymptomatic individuals.

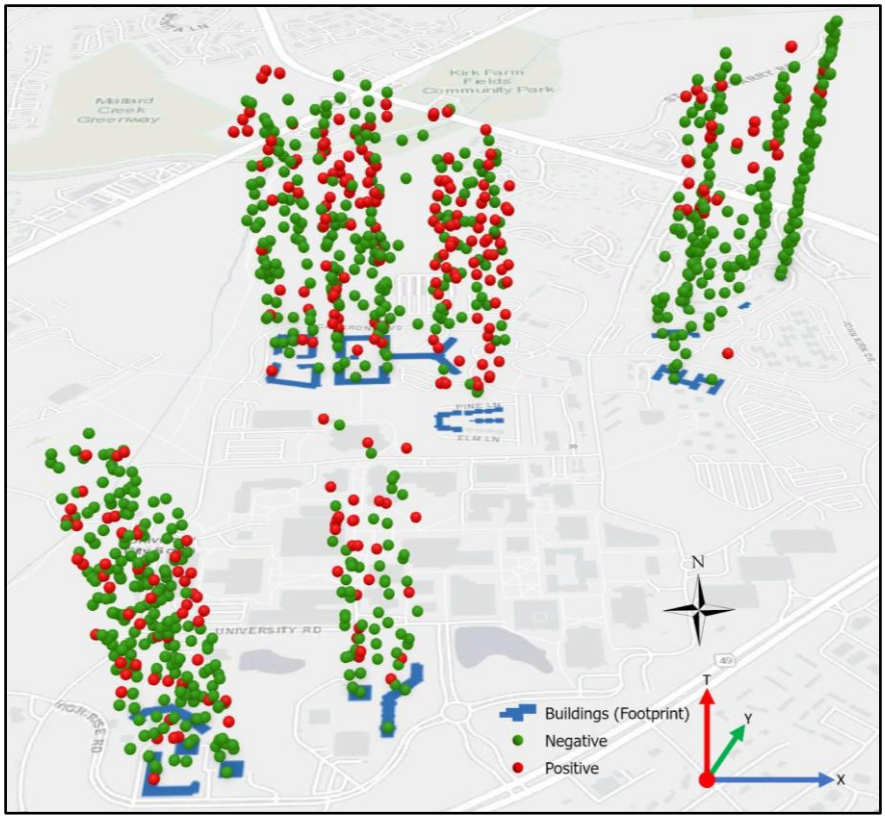

Fig. 7. Illustration of a simulated space-time point pattern of asymptomatic individuals (simulated period: January 4th, 2021 to May 18th, 2021; number of samples: 926; number of positive samples: 264; number of days for perturbation: 3).

3.2.2.3. Similarity analysis of space-time series

To investigate whether any sampling sites show similar responses over time in terms of wastewater testing results, we introduced similarity analysis of time series. We used two similarity metrics, Euclidean distance-based and Dynamic Time Warping (DTW)-based, in this study. Euclidean distance-based metric is a dissimilarity index that evaluates the distance of two time series in the temporal dimension (see Choi et al., 2010). The DTW-based metric allows for comparing time series in terms of shape (see Berndt & Clifford, 1994). DTW is a method that computes the optimal matching between time series (or any sequence patterns) by minimization

of distances (Aghabozorgi et al., 2015; Berndt & Clifford, 1994). Given sampling site *i* and *j*, Euclidean distance-based metric (noted as D_{ij}) between time series of their wastewater testing results can be calculated by Eq. (2). The DTW-based measure (noted as DTW_{ij}) is represented using the shortest cumulative distance between the beginning and end time steps of wastewater testing results at site *i* and *j* once matching between the two time series is optimized (see Eq. 3).

$$D_{ij} = \left(\sum_{k=1}^{m}(w(i,t_k) - w(j,t_k))^2\right)^{1/2} \quad (2)$$

$$DTW_{ij} = C_{ij}(m,m) = d_{ij}(m,m) + \min(C_{ij}(m\text{-}1,m\text{-}1), C_{ij}(m\text{-}1,m), C_{ij}(m,m\text{-}1)) \quad (3)$$

s.t.

$$C_{ij}(0,0)=0;$$

$$d_{ij}(k, l) = |w(i,t_k) - w(j,t_l)|$$

where D_{ij} and DTW_{ij} are the Euclidean distance metric and the dynamic time warping metric of the time series between site *i* and *j*. $w(i,t_k)$ is the binary testing result of sampling site *i* at time t_k, and $w(j,t_l)$ the binary testing result site *j* at time t_l ($k, l \in \{1, 2, ..., m\}$; *m*: number of sampling dates; defined in Eq. 1). $C_{ij}(k,l)$ is the alignment cost between time step t_k of site *i* and time step t_l of site *j*. $d_{ij}(k,l)$ is the distance between time step t_k of site *i* and time step t_l of site *j*. |.| is the absolute function that calculates the absolute distance between site *i* and *j*. min(.) is the function to calculate the minimum of costs. The DTW-based measure is derived using a dynamic programming approach (see Sakoe & Chiba, 1978). Each similarity measure is based on the comparison of two time series, which leads to a *n* by *n* matrix of similarity for our wastewater case (*n*: number of sampling sites; see Eq. 1). Once similarity measures are calculated, hierarchical clustering can be applied to these similarity metrics to compare time series of wastewater testing results of all sampling sites.

3.2.3. Web-based mapping and geovisualization

We used a Web GIS approach (Fu & Sun, 2011; see Peng & Tsou, 2003) for the visual presentation of wastewater data and related spatiotemporal analysis results. Based on the spatiotemporal data model, wastewater data are organized in a spatiotemporal database. We publish these spatially explicit data (sampling sites, sewage network, buildings) into geospatial web services that can be mashed up on a client-side web-based dashboard. When new wastewater testing results are available or the previous sample results are updated, the Web GIS module will automatically update these spatiotemporal data (via API) to the client-side web dashboard (including data, charts, and maps). Further, when new sampling sites are added or some existing sites are retired, the Web GIS module allows for updating spatial data and their geospatial web services (e.g., sampling sites in points, sewersheds in polygons).

We used Esri ArcGIS Online (https://www.arcgis.com/) for Web GIS-based dashboard and ArcGIS API for the automated update of wastewater data to the dashboard. Fig. 8 shows the snapshot of our Web GIS dashboard. The web mapping interface shows the locations of buildings, samplers, and sewersheds (aka, neighborhoods), and sewer networks (hidden for confidentiality consideration). Moreover, the color scheme of samplers and buildings indicate the sample testing results (shown in the map legend). Summary of wastewater testing results including number of positive buildings, sampling sites, sewershed sites, and their time series is displayed (for example, in charts). This provides visual and interactive analytics support that can inform decision makers for subsequent decision making on, for example, clinical testing or contract tracing.

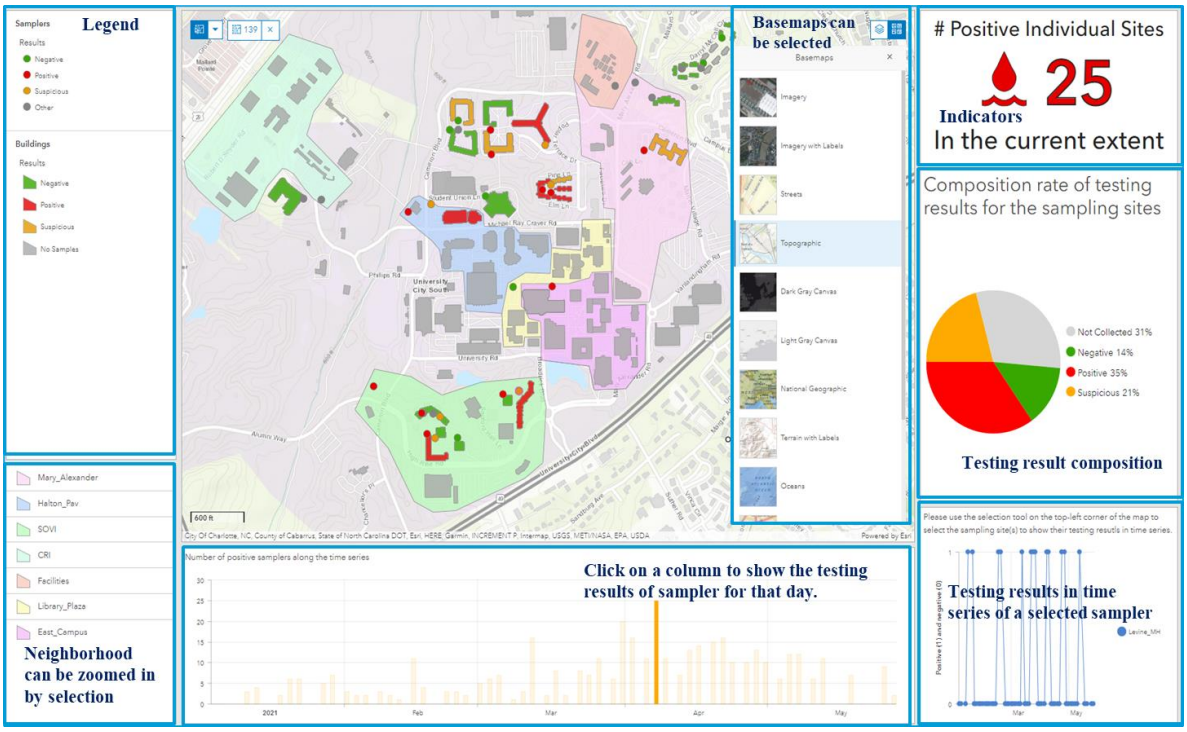

Fig. 8. Snapshot of the Web GIS dashboard (sewer networks is hidden due to confidentiality consideration).

3.2.4. Implementation

Our web-based SDSS is implemented within a web server. Jupyter Notebooks (https://jupyter.org/) were used to implement the web-based main interface of the SDSS and access to its individual modules. Table 2 shows the software or libraries used to implement each individual module of the SDSS. We use ArcGIS API for Python to update wastewater testing results to the Web GIS dashboard based on ArcGIS Online. Google OAuth was chosen as the authentication mechanism of a web-based system for automated data synchronization.

467 **Table 2.** Software or libraries used by the web-based SDSS for wastewater surveillance.

Module name	Sub-module	Software/Library	URL
Module for geospatial database design and data synchronization	Web interface for data synchronization	ArcGIS API for Python (v1.9.1) Flask (v3.1)	https://developers.arcgis.com/python/ https://flask.palletsprojects.com
Module for spatiotemporal analysis	Space-time cluster detection	SatScan	https://www.satscan.org/
	Space-time simulation of point-type events	Python scripts	n/a
	Similarity measures of time series	TSdist v3.1 - Distance Measures for Time Series in R	https://cran.r-project.org/web/packages/TSdist/index.html
Module for web-based mapping and geovisualization	Web GIS dashboard	Esri ArcGIS Online	https://www.arcgis.com/index.html

468

469 ## 4. Results

470 ### 4.1. Overall results

471 Our wastewater surveillance initiative has been collecting wastewater data since Fall 2020. We
472 have established 38 sampling sites since then. These sites provide strong support for monitoring
473 the COVID-19 situation via the wastewater surveillance approach. Wastewater testing results are
474 uploaded, synchronized, processed, analyzed, and visualized via the web-based SDSS. In this
475 study, we focus on using wastewater testing results from 23 residence hall sites from
476 01/04/2021to 05/18/2021 (in total 135 days) as we have consistently used these sites to collect
477 samples during this period (results for neighborhoods sites and sampling sites of residence halls
478 that were established or removed during this period were excluded). Sewage samples were
479 collected three times a week on Monday, Wednesday, and Fridays for Spring 2021. This leads to

54 sample collections for each sampling site during the study period (18 weeks times 3 collections per week). However, it is not always possible to collect a sample at every site every time due to variations in flow or unexpected physical obstruction of the autosampler probe. As a result, 926 samples were collected from these 23 sites for Spring 2021. Among them, there are 662 negative (71.49%), and 264 positive (28.50%).

Fig. 9 depicts the number of positive sampling sites during the study period compared to the 7-day averaged number of cases in Mecklenburg County, NC (original data is retrieved from the U.S. Centers for Disease Control and Prevention, https://ephtracking.cdc.gov/DataExplorer/). As we could see, the number of positive sites fluctuates between 0 and 8 before March 24th, 2021. After that date, an increasing pattern in terms of the number of positive sampling sites can be observed and lasts for about 2 weeks. This number reaches its maximum (16) on April 9th, 2021. After April 9th, the number drops to under 10 and tends to show a decreasing pattern over time.

The spring semester of the University was postponed to start from January 20th, 2021 and Spring Break was changed to the week from February 8th to 13th, which was a decision made by the university due to the consideration of the pandemics (number of cases in Mecklenburg County is high in January and February; see Fig. 9). This explains the lower number of positive wastewater samples during the early stage of the semester. An increase in the number of cases in Mecklenburg County (corresponding to the local peak of the SARS-CoV-2 Alpha variant) appeared from mid March to mid April, 2021. Relaxation of local COVID-19 restrictions may also have contributed to this peak (see Executive Orders No. 195 and No. 204 by the North Carolina Governor on February 26th (NC government, 2021a) and March 26th (NC government, 2021b)). This corresponds to (and may explain) a dramatic increase in the number of positive

samples on campus during that period. Decreasing trends appeared from mid to late April, 2021 in Mecklenburg County in terms of number of cases and on campus with respect to the number of positive wastewater samples. This can be attributed to the availability of vaccines to more people (increase in vaccination rate). Students started to receive vaccines beginning on March 31st, 2021, and vaccines were available to all adults in North Carolina by April 7th (Source: https://governor.nc.gov). Two on-campus vaccine clinics (March 31st, 2021, and April 12th, 2021) hosted by the university facilitated vaccine uptake by students and faculty. All of these vaccine-related events play an important role in contributing to the decreasing number of positive samples in the final weeks of the semester.

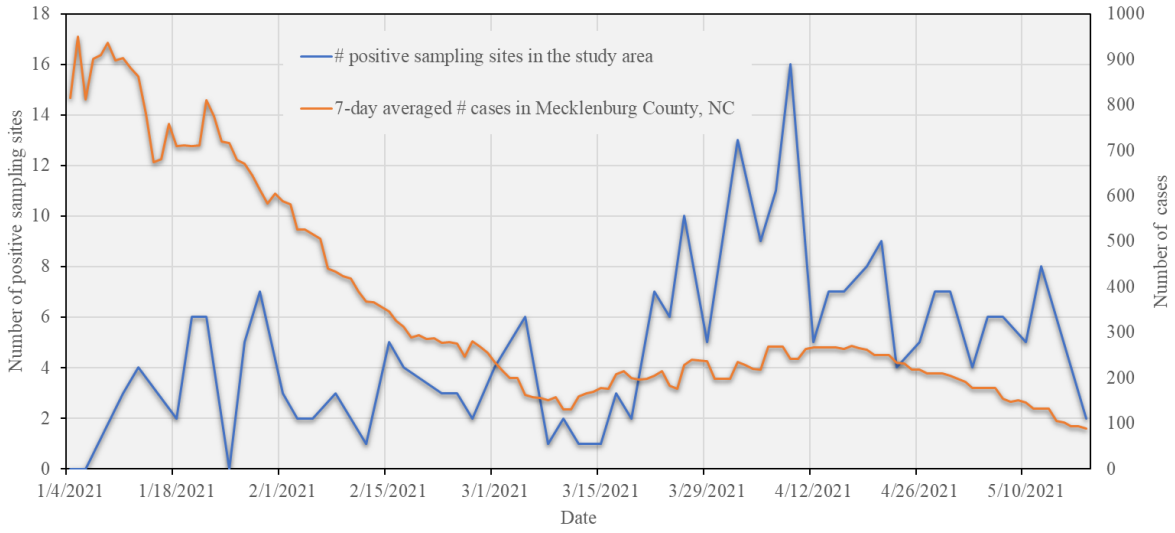

Fig. 9. Number of positive sampling sites in the study area and 7-day averaged number of cases in Mecklenburg County, NC over the study period.

4.2. Results of space-time cluster analysis

The use of space-time scan statistics needs to determine the upper limit of the spatiotemporal cluster size bandwidth (spatial bandwidth and temporal bandwidth). For the upper limit of the spatial bandwidth, we set the maximum spatial cluster parameter (corresponding to the

518 percentage of population at risk—i.e., number of collected samples for this study) as 50%. The
519 upper limit of the temporal bandwidth is set to 50% of the duration of the study period.

520 **4.2.1. Space-time cluster analysis results based on samples collected from sampling sites**

521 Fig. 10 and Table 3 depict the space-time scan results based on the collected samples for which
522 the locations of samplers were used as coordinates for space-time scan analysis. Fig. 10 shows
523 the map of detected space-time clusters. One significant cluster (p-value under 5%; based on 999
524 Monte Carlo runs) was detected that contains two sampling sites lasting from March 17th, 2021
525 to April 30th, 2021 (in total 44 days--about 7 weeks). These two sampling sites cover three
526 residence halls. Both the total number of collected samples (population for space-time scan) and
527 number of positive samples (cases) are 34, indicating all collected samples are positive in the
528 detected space-time cluster during these 7 weeks. The detection of this significant cluster is
529 because the three buildings have been used by the University for isolation and quarantine
530 purposes. The relative risk is 3.88 in the detected cluster, indicating the residence halls covered
531 by the sampling sites within the clusters are around 3-4 times higher than those out of the
532 clusters in terms of the ratio of number of positive samples over expected value.

533 Table 3. Information of the detected space-time cluster based on locations of sampling sites.

Parameter	Value	Description
Time span	3/17/2021 to 4/30/2021	Start date and end date of the cluster
Population	34	Number of collected samples
Number of cases	34	Number of positive samples
Expected cases	9.69	The number of samples within the cluster multiplied by the ratio of the total number of positive samples over the total number of samples for the entire study region.
Estimated risk	3.51	The ratio of the number of positive samples within the cluster over the number of expected cases within the cluster
Relative risk	3.88	The ratio of the estimated risk within the cluster over that outside of the cluster
p-value	5.2E-15 (p<=0.05)	p-value based on 1,000 Monte Carlo runs

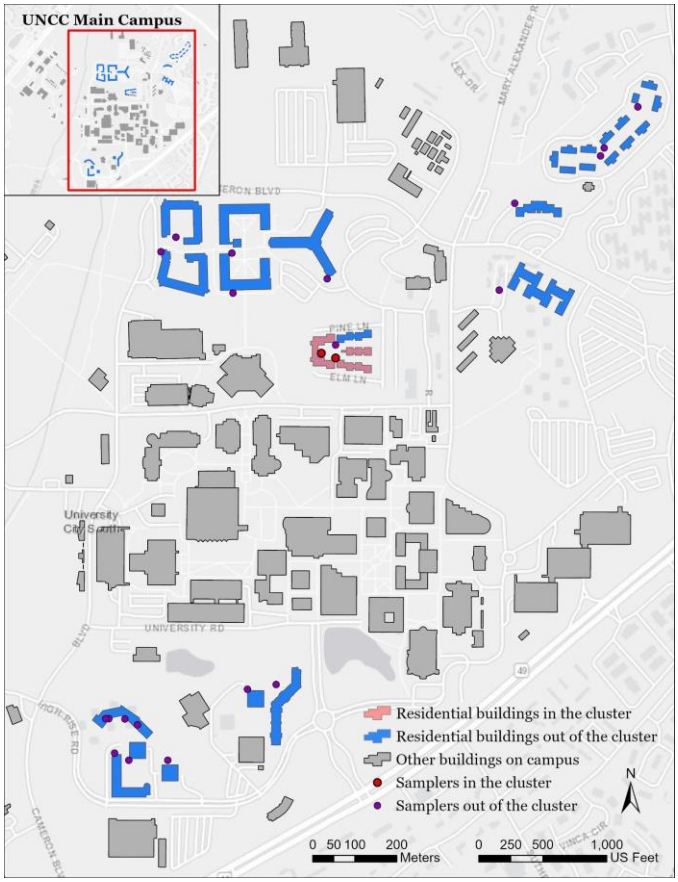

Fig. 10. Map of the sampling sites in the detected cluster and corresponding residence halls (sewer networks were hidden due to confidentiality consideration)

4.2.2. Space-time cluster analysis results from simulated space-time point patterns

The space-time scan results using locations of collected samples are based on sampling sites. In our case study, these wastewater samples were contributed from individuals living in their residence halls. Our sampling sites are, however, either outside or inside of residence halls, thus posing an issue of locational uncertainty. To address this issue, we used the space-time simulation of point-type events. We associate the binary (positive/negative) wastewater sampling results from sampling sites back to the residence halls. For those sites that cover a single building, once the wastewater testing result from any of these sites is positive, the residence hall will exhibit a positive signal. For a sampling site that covers multiple buildings, all these covered

buildings will be positive if the testing results from the site are positive. The number of simulations for generating space-time point patterns was set to 1,000 in this study.

Fig. 11 shows the spatial pattern of residence halls within the detected clusters from 1,000 simulated space-time point patterns (if a simulated presymptomatic individual within a building belongs to a cluster, then we consider the building is within the cluster). There are 8 residence halls that are within significant space-time clusters (at a 95% confidence level). We hide the names of the residence halls for confidentiality purposes. Table 4 summarizes the information of detected clusters based on the 1,000 simulated space-time point patterns. Relative risk within clusters is 2.774, indicating the estimated risk of residence halls within the cluster is 2-3 times higher than that outside the cluster.

Table 5 depicts start and end dates of each building within clusters, and Table 6 illustrates the number of weeks that the detected space-time clusters from 1,000 simulations last. Fig. 12 shows the histogram of the number of clusters in terms of the start date and end date of a building within detected clusters. It can be observed from Table 6 that detected clusters last from 1 week to 6 weeks, and 92.8% of the clusters last around 3-5 weeks. In general, the significant start date of clusters on each building at high risk concentrates on March 24th, 2021 (one exception is March 26th for building 7) and most of them end around April 23rd or 24th (April 20th for building 7), lasting around 1 month. This suggests that the wastewater signals from these 8 buildings correspond to the second peak of the pandemic in Mecklenburg County (see Fig. 9). The three buildings used for isolation and quarantine purposes are included in these 8 buildings.

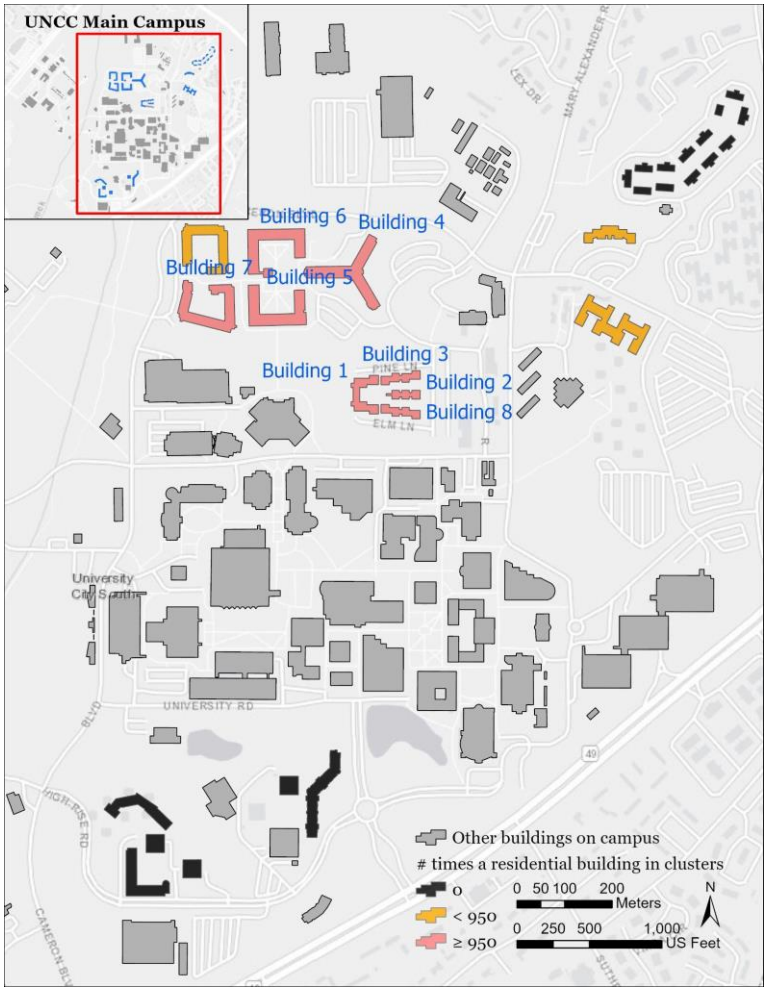

Fig. 11. Map of the residence halls in the detected clusters from simulated space-time point patterns (number of simulations: 1,000)

Table 4. Summary of the clusters detected in 1,000 simulated datasets.

	Mean	Standard Deviation	Minimum	Maximum	Confidence Level for mean (95%)
Population	156.144	26.971	32	205	1.674
Number of cases	94.292	11.547	32	112	0.717
Expected cases	44.516	7.689	9.123	58.445	0.477
Estimated risk	2.145	0.167	1.901	3.508	0.010
Relative risk	2.774	0.129	2.544	3.904	0.008
P-value	1.08E-11	2.21E-11	6.99E-15	2.26E-10	1.37E-12

574 Table 5. Start and end dates of the buildings detected within clusters based on 1,000 simulations.

Building index	Start date (p<=0.05)	End date (p<=0.05)	Number of days at high risk
Building 1	March 23rd, 2021	April 24th, 2021	33
Building 2	March 23rd, 2021	April 24th, 2021	33
Building 3	March 23rd, 2021	April 24th, 2021	33
Building 4	March 23rd, 2021	April 23rd, 2021	32
Building 5	March 23rd, 2021	April 23rd, 2021	32
Building 6	March 23rd, 2021	April 23rd, 2021	32
Building 7	March 26th, 2021	April 20th, 2021	26
Building 8	March 23rd, 2021	April 24th, 2021	33

575

576 Table 6. Number of weeks covered by the space-time clusters detected from simulated patterns
577 (number of simulations: 1,000).

#Weeks	Frequency
1 week	53
2 weeks	18
3 weeks	254
4 weeks	384
5 weeks	290
6 weeks	1

578

579

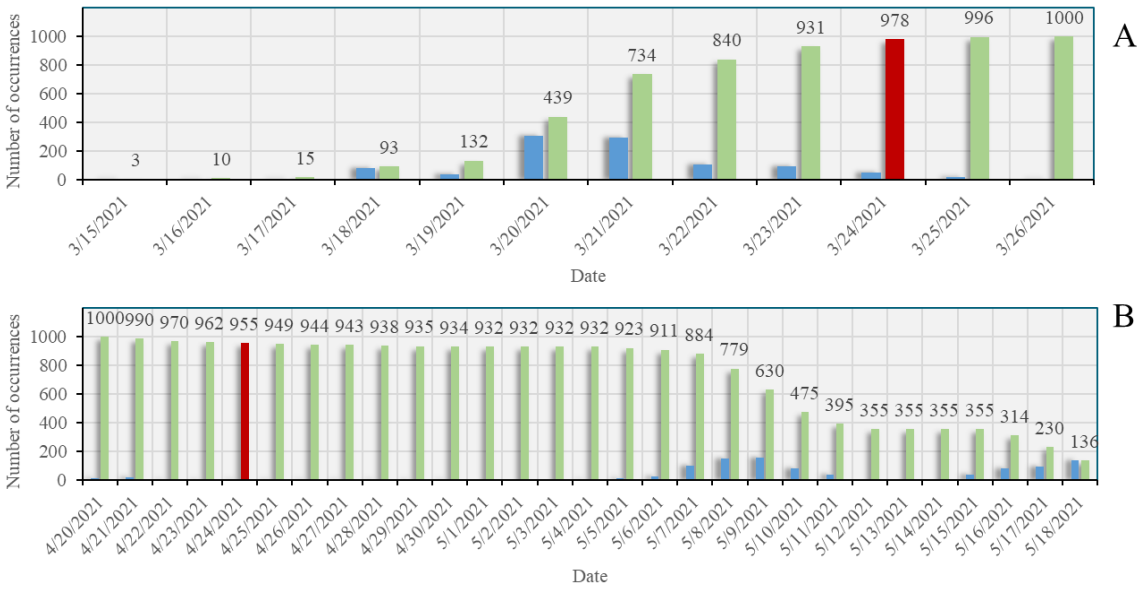

Fig. 12. Histograms of the start (A) and end (B) dates that a building (Building 1) was identified as within a cluster (blue) and the number of occurrences that a building was identified as within a cluster over time (green) from simulations. Significant start and end dates (95% confidence level) were colored in red. Number of simulations: 1,000.

The use of space-time scan for cluster analysis is computationally demanding because each analysis would need additional 999 Monte Carlo runs for significance testing, and we need to conduct this analysis on 1,000 simulated space-time point patterns of presymptomatic individuals. To address this computational challenge, we deployed these analyses to a high performance computing (HPC) cluster (computing node configuration: dual 24-Core Intel Xeon Gold 6248R CPU with clock rate of 3.00 GHz and 384GB memory). Twenty computing nodes (each with 24 cores--i.e., in total 480 CPUs) were used for acceleration. The parallel computing time of the analysis of a single simulated point pattern on a computing node varies from 7.76 to 16.26 minutes with a mean of 8.42 minutes, while the mean sequential computing time for a single analysis is 139.36 minutes. The total parallel computing time on 480 CPUs for 1,000 analyses is 7.08 hours, compared with the total sequential computing time (on a single CPU) of

2,322.72 hours (around 97 days). As a result, 327.91 times of acceleration was achieved for these analyses by using 480 CPUs.

4.3. Results of Similarity Analysis of Time Series

We conducted similarity analyses based on the time series of wastewater testing results from the 23 sampling sites over the study period. Fig. 13 shows the results of similarity analysis with respect to metrics of Euclidean distance and DTW. Both Euclidean distance and DTW are dissimilarity metrics, meaning that the larger the value of the metrics, the more dissimilar the time series of two sites are. We then applied hierarchical clustering analysis to each of the two metrics. Elbow method (Thorndike, 1953) was used to determine the number of clusters based on these metrics. As a result, two clusters were identified with respect to the Euclidean distance metric and three clusters for the DTW metric.

Fig. 14 depicts the cluster dendrograms of the two similarity metrics as well as the spatial distribution of the identified clusters with respect to each of the similarity metrics. Fig. 15 shows the number of positive sites per week for each group identified by similarity metrics. The Euclidean distance-based metric clusters the sampling sites to two groups, whereas there are three main groups identified by the DTW metric. In terms of Euclidean distance-based metric (see Fig. 15A), group 1 covers five sampling sites, about 22% over 23 sampling sites in this study. The number of positive samples of group 1 fluctuates around 5 positive samples per collection day before and on March 15th, 2021. It rises to 10 - 15 positive samples per collection day from late March to mid April and then a decreasing trend appears until mid May. Group 2 has 18 sampling sites, around three times higher than those in group 1. The number of positive

samples for each group in the study period tends to be close compared with the total number of sampling sites in each group, indicating that buildings in group 1 are at higher risk of being exposed under virus than those in group 2. We can also observe a rising pattern in the number of positive samples for group 2 in mid March and a decreasing trend from late April to the end of the study period.

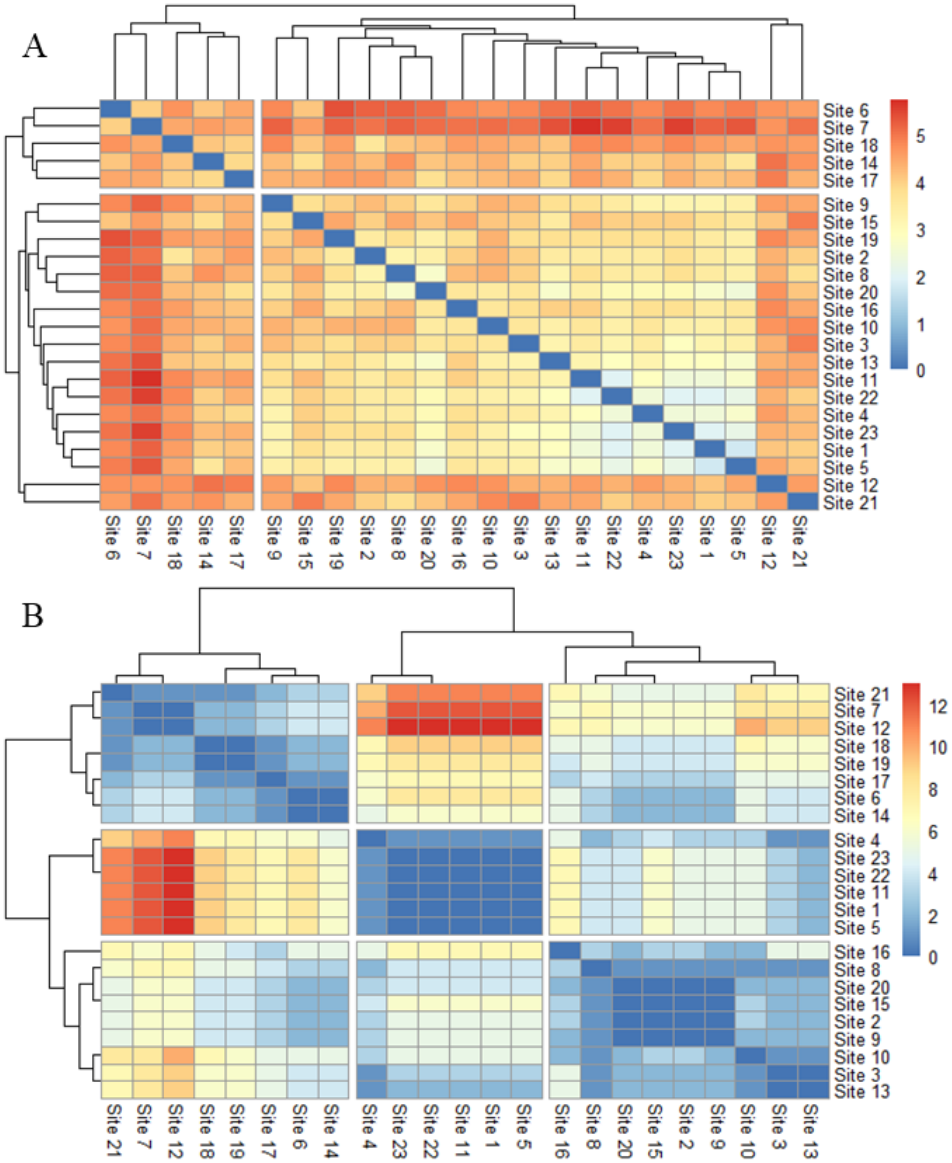

Fig. 13. Matrix of similarity metrics. (A: Euclidean distance; B: Dynamic Time Warping).

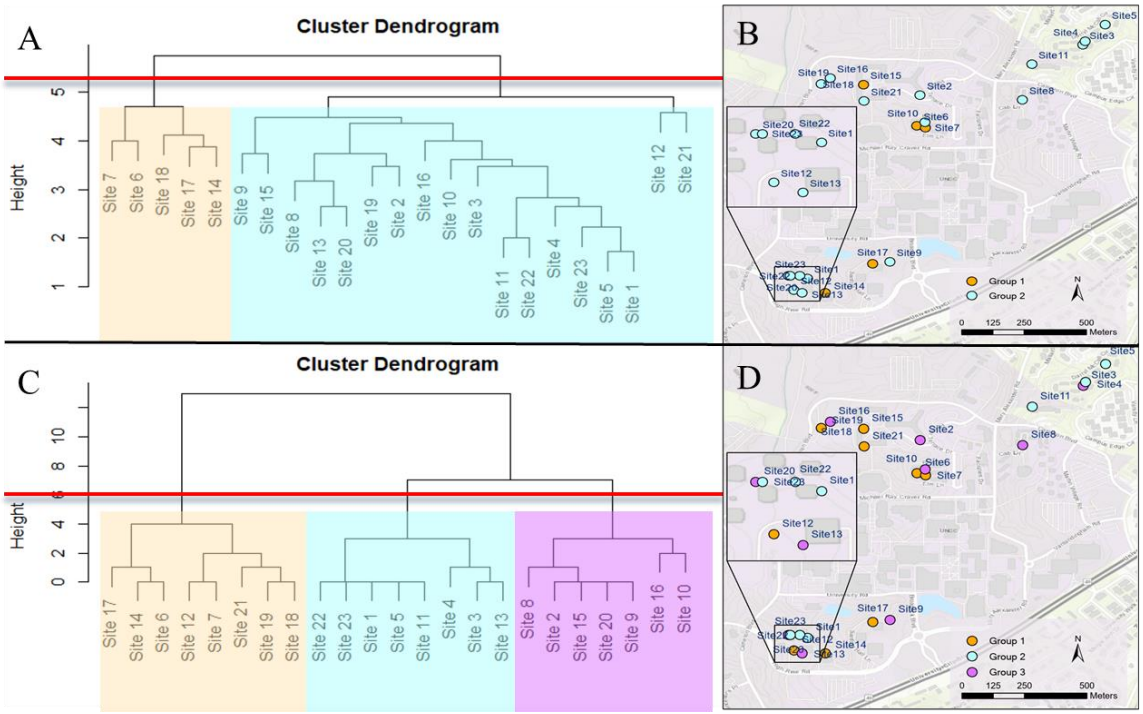

Fig. 14. Cluster dendrograms of similarity metrics and spatial patterns of clustered results. A and B are for Euclidean distance metric. C and D are for Dynamic Time Warping metric. The cut-off of the number of clusters (red line) was identified using the Elbow curves. Group 1, 2, and 3 were shaded in orange, blue, and purple.

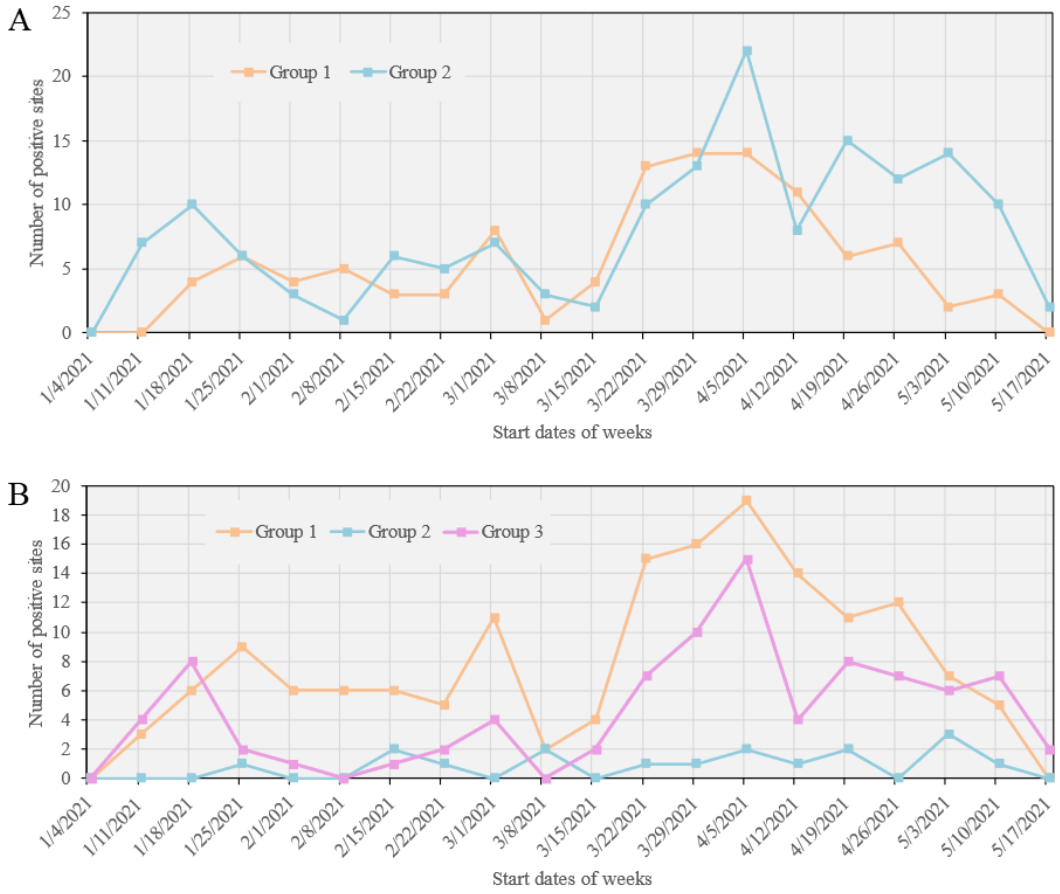

Fig. 15. Number of positive sampling sites per week for each group identified by similarity metric over time. A: for the Euclidean distance metric. B: for the Dynamic Time Warping metric. The horizontal axis shows the start date of each week. The last week starting from May 17th only has two-days data available.

With respect to the DTW metric (see Fig. 15B), three groups are identified, where the number of sampling sites are 8, 6, and 9 for group 1, 2, and 3. It is observed that the weekly number of positive samples in group 1 is higher than those of groups 2 and 3 in between January 25th, 2021 and May 3rd, 2021, covering most of the study period. The number of weekly positive samples in group 2 is higher than that in group 3 especially in the beginning of the study period until February 8th, 2021, and from March 15th to May 18th, 2021. Group 3 stays between 0 to 3 positive samples per week during this time span. Group 1 and 2 exhibit similar responses to the

spread of COVID-19 as we can observe three peaks in the time series: around January 20th, March 1st, and April 5th. Both group 1 and 2 strongly responded to the wave in Mecklenburg County, NC starting from mid March, 2021 (see Fig. 9); however, group 3 did not show a significant reaction to this wave, indicating residence halls in this group appear a relatively lower risk of being exposed to the virus than others during the study period.

Sites 6 and 7 identified in group 1 of both similarity metrics (see Fig. 14) are also detected within the cluster using space-time scan (see Fig. 10), indicating that buildings related to the two sites are more likely to be under exposure of the COVID-19 during the study period. Site 14, 17, and 18 in group 1 for Euclidean distance are also included in the group 1 of DTW metric, indicating these sites also need to be paid attention. Further, group 1 of DTW metric suggests that site 12, 19, and 21 are at relatively higher risk as well. Buildings in group 2 of the Euclidean distance metric appear less likely to be under exposure of the virus than those in group 1. It can be observed that buildings in group 2 and 3 detected by the DTW metric are included in group 2 of the Euclidean distance. Results in Fig. 15B also suggests that the two groups of DTW metric, especially group 2, appear to be characterized by a relatively low number of positive wastewater samples during the study period.

5. Discussion

Our web-based SDSS provides support for automating data operations, analysis and modeling, and visualization capabilities within an integrative environment. Wastewater surveillance is dependent on various data that may cut across different spatiotemporal scales. Our web-based SDSS allows for automated synchronization and mapping of these spatiotemporal data. This

provides timely support for the early detection of the COVID-19 virus in campus wastewater and thus greatly facilitates the monitoring and mitigation of the pandemic situation in the University. At the same time, the management of space-time wastewater data within this integrated environment can help monitor the status of samplers and their sampling sites. If any issues occur to the autosamplers that lead to the unavailability of samples over time, we could quickly identify and resolve the issues with support from this SDSS, thus ensuring the continual functioning of samplers.

Wastewater surveillance data are spatiotemporally explicit. Spatiotemporal analysis and modeling can be of great help in discovering interesting patterns in these spatiotemporal data, represented by the clusters of positive samples or residence halls detected using space-time scan approach and similarity analysis of space-time series in this study. The combination of the spatiotemporal analysis approaches has been suggested in the literature (see Xu & Beard, 2021). Space-time scan methods, represented by SatScan in this study, allows for detecting the co-occurrence of space-time events (positive samples in this study) within a specific time period (i.e., local- or regional-level analysis). Further, similarity analysis of space-time series offers a means of comparing space-time events over the entire study period--i.e., system-level comparison. Combining these spatiotemporal analysis methods enables us to discover patterns of interest from different levels (with respect to the study system of interest). On the one hand, this combined approach allows for identifying those residence halls where interactions with their residents are at a high risk during specific time periods. On the other hand, it gives us recommendations on the group of residence halls with a lower risk of virus even when it was peaking. This combined analysis approach provides substantial support for addressing

688 spatiotemporal questions (as in the Introduction section). It is also noted that the detection of
689 these space-time clusters may be biased as samples may not be collected from every site each
690 time, which will be investigated in future work. However, in general, these detected clusters
691 from spatiotemporal analysis and modeling provide invaluable and critical support for the
692 University on decisions or guidelines for the prevention of outbreak of the virus and control of
693 virus transmission on campus.

694 The use of the space-time simulation of presymptomatic individuals was necessary because the
695 relationship between sampling sites and their associated buildings is complicated (instead of one-
696 to-one mapping) and because individuals in residence halls are sources that contribute to the
697 wastewater testing results instead of samplers at sampling sites. The space-time scan results
698 based on simulated individuals in residence halls are different than those based on sampling
699 sites. The former approach detects more residence halls within the clusters of positive
700 wastewater samples. The space-time simulation of the presymptomatic individuals provides an
701 alternative approach for the possible locations of these individuals instead of relying on the
702 sampling sites. While the detected clusters include more residence halls from space-time
703 simulation, it is better than underestimating the number of residence halls that may exhibit strong
704 positive signals of COVID-19 virus in wastewater. Of course, these clusters of positive
705 wastewater samples are based on the space-time scan, which is a statistically based exploratory
706 data analysis approach. The further interpretation of these clusters would require the expert
707 knowledge from the collaboration of domain scientists (e.g., biogenetic professionals), better
708 understanding of the wastewater surveillance system (e.g., sampling sites, residence halls,
709 student interactions), and the incorporation of clinical testing and contact tracing data. In

particular, clinical testing data could be used to further improve the space-time simulation of presymptomatic individuals in terms of model calibration and validation. For example, in this study, wastewater samples that have 2 or less replicates producing signals are treated as negative. The use of clinical testing data could help us to fine tune the relationship between wastewater signals and infected individuals for more reliable spatiotemporal cluster analysis.

Web-based GIS is of essence in this web-based SDSS in terms of visual presentation of space-time data related to wastewater surveillance. Web-based GIS technologies and geospatial web services have been increasingly developed and available for the online management and mapping of spatially explicit data. However, the automatic update of data to Web GIS dashboards has been the bottleneck of Web GIS applications. Our web-based GIS and visualization module provides automation support that allows for the automatic update of wastewater sample data to the web GIS dashboard. Specifically, we aimed to reduce the time and number of steps that data are taken from the lab to the dashboard. This module will lead to the saving of tremendous time and cost as required by the update and dissemination of wastewater data that are continuously available over time.

6. Conclusions

The web-based SDSS framework presented in this study empowers the management, analytics and sharing of wastewater surveillance-related data at multiple spatiotemporal scales. The SDSS framework serves as a synergistic platform that integrates various types of data based on the spatiotemporal data model. Spatiotemporal analysis and modeling capabilities incorporated in this framework offer a means of unveiling interesting or unexpected patterns from the

731 wastewater data. These patterns may not be easily detected using visual inspection. These data-
732 and model-related capabilities are managed and automated within the SDSS framework to ensure
733 their reusability and the reproducibility of analytic results. This SDSS framework, built in with
734 Web GIS dashboard functionality, will inform critical decision-making and guideline
735 development for monitoring COVID-19 situations in the study region.

736 Future work of our study includes: 1) integration of 3D BIM-based building model into the web-
737 based environment, 2) adding more spatial modeling capabilities (e.g., spatial simulation for
738 scenario analysis and representation of individual behavior and social behavior using agent-
739 based modeling; spatial optimization for optimal allocation of sampling sites), 3) use of
740 continuous variable of virus concentration in wastewater samples instead of binary indicator for
741 spatiotemporal analysis, and 4) extend the web-based SDSS framework to other or larger regions
742 by, for example, linking to city sewage network and wastewater treatment plants at regional
743 level.

Acknowledgements

The authors would like to thank Chancellor Sharon Gaber, Provost Joan Lorden, and Richard Tankersley, Vice Chancellor for Research and Economic Development and his team for strong institutional support of this wastewater surveillance project, Deborah Thomas, Chair of the Department of Geography and Earth Sciences for facilitating setting up geospatial computing needs for the project. The authors owe thanks to Facilities Management (Greg Cole) and OneIT (Alex Chapin, Elie Saliba) at the University of North Carolina at Charlotte for their support and help on sampling site setup and computing needs. High-performance computing resources used in this project were provided by University Research Computing at the University of North Carolina at Charlotte.

Funding:

The authors would like to thank financial support through CARES funding from NC General Assembly and funding from the University of North Carolina at Charlotte.

References

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering–a decade review. *Information Systems*, *53*, 16–38. https://doi.org/10.1016/j.is.2015.04.007

Ahmed, W., Tscharke, B., Bertsch, P. M., Bibby, K., Bivins, A., Choi, P., Clarke, L., Dwyer, J., Edson, J., Nguyen, T. M. H., O'Brien, J. W., Simpson, S. L., Sherman, P., Thomas, K. V., Verhagen, R., Zaugg, J., & Mueller, J. F. (2021). SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: A temporal case study. *The Science of the Total Environment*, *761*, 144216. https://doi.org/10.1016/j.scitotenv.2020.144216

Armstrong, M. P., Densham, P. J., & Rushton, G. (1986). Architecture for a microcomputer based spatial decision support system. *Second International Symposium on Spatial Data Handling, International Geographic Union*, 120–131. https://iro.uiowa.edu/esploro/outputs/conferenceProceeding/Architecture-for-a-microcomputer-based-spatial/9983557551002771

Barua, V. B., Juel, M. A. I., Blackwood, A. D., Clerkin, T., Ciesielski, M., Sorinolu, A. J., Holcomb, D. A., Young, I., Kimble, G., Sypolt, S., Engel, L. S., Noble, R. T., & Munir, M. (2021). Tracking the temporal variation of COVID-19 surges through wastewater-based epidemiology during the peak of the pandemic: A six-month long study in Charlotte, North Carolina. *The Science of the Total Environment*, 152503. https://doi.org/10.1016/j.scitotenv.2021.152503

Becerik-Gerber, B., Jazizadeh, F., Li, N., & Calis, G. (2012). Application Areas and Data Requirements for BIM-Enabled Facilities Management. *Journal of Construction*

Engineering and Management, *138*(3), 431–442. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000433

Berndt, D. J., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *KDD Workshop*, *10*, 359–370. https://www.aaai.org/Library/Workshops/1994/ws94-03-031.php

Bowes, D. A., Driver, E. M., Kraberger, S., Fontenele, R. S., Holland, L. A., Wright, J., Johnston, B., Savic, S., Newell, M. E., & Adhikari, S. (2021). Unrestricted Online Sharing of High-frequency, High-resolution Data on SARS-CoV-2 in Wastewater to Inform the COVID-19 Public Health Response in Greater Tempe, Arizona. *medRxiv*. https://doi.org/10.1101/2021.07.29.21261338

Chen, B. Y., Yuan, H., Li, Q., Shaw, S.-L., Lam, W. H. K., & Chen, X. (2016). Spatiotemporal data model for network time geographic analysis in the era of big data. *International Journal of Geographical Information Science*, *30*(6), 1041–1071. https://doi.org/10.1080/13658816.2015.1104317

Choi, S. S., Cha, S. H., & Tappert, C. C. (2010). A survey of binary similarity and distance measures. *Journal of Systemics, Cybernetics and Informatics*, *8*(1), 43–48. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.352.6123&rep=rep1&type=pdf

Ciesielski, M., Blackwood, D., Clerkin, T., Gonzalez, R., Thompson, H., Larson, A., & Noble, R. (2021). Assessing sensitivity and reproducibility of RT-ddPCR and RT-qPCR for the quantification of SARS-CoV-2 in wastewater. *Journal of Virological Methods*, *297*, 114230. https://doi.org/10.1016/j.jviromet.2021.114230

Crimi, A., Jones, T., & Sgalambro, A. (2019). Designing a Web Spatial Decision Support

System Based on Analytic Network Process to Locate a Freight Lorry Parking. *Sustainability: Science Practice and Policy*, *11*(20), 5629. https://doi.org/10.3390/su11205629

Delmelle, E., Zhu, H., Tang, W., & Casas, I. (2014). A web-based geospatial toolkit for the monitoring of dengue fever. *Applied Geography* , *52*, 144–152. https://doi.org/10.1016/j.apgeog.2014.05.007

Densham, P. J. (1991). Spatial decision support systems. In D. J. Maguire, M. F. Goodchild, & D. W. Rhind (Eds.), *Geographical Information Systems: Principles and Applications* (Vol. 1, pp. 403–412). Hoboken, NJ: John Wiley & Sons.

Desjardins, M. R., Hohl, A., & Delmelle, E. (2020). Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied Geography* , *118*, 102202. https://doi.org/10.1016/j.apgeog.2020.102202

Diggle, P. J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition*. CRC Press.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, *20*(5), 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1

Fox, M. D., Bailey, D. C., Seamon, M. D., & Miranda, M. L. (2021). Response to a COVID-19 outbreak on a University Campus—Indiana, August 2020. *Morbidity and Mortality Weekly Report*, *70*(4), 118. https://doi.org/10.15585/mmwr.mm7004a3

Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F., & Billa, L. (2020). Spatial analysis and

GIS in the study of COVID-19. A review. *The Science of the Total Environment*, *739*, 140033. https://doi.org/10.1016/j.scitotenv.2020.140033

Fu, P., & Sun, J. (2011). *Web GIS: Principles and Applications*. Redlands, CA: Esri Press.

Ghosh, D. (2008). A loose coupling technique for integrating GIS and multi-criteria decision making. *Transactions in GIS*, *12*(3), 365–375. https://doi.org/10.1111/j.1467-9671.2008.01103.x

Gibas, C., Lambirth, K., Mittal, N., Juel, M. A. I., Barua, V. B., Roppolo Brazell, L., Hinton, K., Lontai, J., Stark, N., Young, I., Quach, C., Russ, M., Kauer, J., Nicolosi, B., Chen, D., Akella, S., Tang, W., Schlueter, J., & Munir, M. (2021). Implementing building-level SARS-CoV-2 wastewater surveillance on a university campus. *The Science of the Total Environment*, *782*, 146749. https://doi.org/10.1016/j.scitotenv.2021.146749

Harris-Lovett, S., Nelson, K. L., Beamer, P., Bischel, H. N., Bivins, A., Bruder, A., Butler, C., Camenisch, T. D., De Long, S. K., Karthikeyan, S., Larsen, D. A., Meierdiercks, K., Mouser, P. J., Pagsuyoin, S., Prasek, S. M., Radniecki, T. S., Ram, J. L., Roper, D. K., Safford, H., … Korfmacher, K. S. (2021). Wastewater Surveillance for SARS-CoV-2 on College Campuses: Initial Efforts, Lessons Learned, and Research Needs. *International Journal of Environmental Research and Public Health*, *18*(9), 4455. https://doi.org/10.3390/ijerph18094455

Hohl, A., Delmelle, E., Desjardins, M. R., & Lan, Y. (2020). Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spatial and Spatio-Temporal Epidemiology*, *34*, 100354. https://doi.org/10.1016/j.sste.2020.100354

Juel, M. A. I., Stark, N., Nicolosi, B., Lontai, J., Lambirth, K., Schlueter, J., Gibas, C., & Munir,

M. (2021). Performance evaluation of virus concentration methods for implementing SARS-CoV-2 wastewater based epidemiology emphasizing quick data turnaround. *The Science of the Total Environment*, *801*, 149656. https://doi.org/10.1016/j.scitotenv.2021.149656

Karthikeyan, S., Nguyen, A., McDonald, D., Zong, Y., Ronquillo, N., Ren, J., Zou, J., Farmer, S., Humphrey, G., Henderson, D., Javidi, T., Messer, K., Anderson, C., Schooley, R., Martin, N. K., & Knight, R. (2021). Rapid, Large-Scale Wastewater Surveillance and Automated Reporting System Enable Early Detection of Nearly 85% of COVID-19 Cases on a University Campus. *mSystems*. https://doi.org/10.1128/mSystems.00793-21

Keenan, P. B., & Jankowski, P. (2019). Spatial Decision Support Systems: Three decades on. *Decision Support Systems*, *116*, 64–76. https://doi.org/10.1016/j.dss.2018.10.010

Kim, S., & Castro, M. C. (2020). Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). *International Journal of Infectious Diseases*, *98*, 328–333. https://doi.org/10.1016/j.ijid.2020.07.004

Kırbıyık, U., Binder, A. M., Ghinai, I., & Zawitz, C. (2020). Network Characteristics and Visualization of COVID-19 Outbreak in a Large Detention Facility in the United States—Cook County, Illinois, 2020. *Morbidity and Mortality Weekly Report. Surveillance Summaries* , *69*(44), 1625. https://doi.org/10.15585/mmwr.mm6944a3

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods*, *26*(6), 1481–1496. https://doi.org/10.1080/03610929708831995

Kulldorff, M. (1999). Spatial scan statistics: Models, calculations, and applications. In *Scan Statistics and Applications* (pp. 303–322). Birkhäuser Boston. https://doi.org/10.1007/978-

1-4612-1578-3_14

Kulldorff, M., Athas, W. F., Feurer, E. J., Miller, B. A., & Key, C. R. (1998). Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health*, *88*(9), 1377–1380. https://doi.org/10.2105/ajph.88.9.1377

Lam-Hine, T., McCurdy, S. A., Santora, L., Duncan, L., Corbett-Detig, R., Kapusinszky, B., & Willis, M. (2021). Outbreak associated with SARS-CoV-2 B. 1.617. 2 (delta) variant in an elementary school—Marin County, California, May–June 2021. *Morbidity and Mortality Weekly Report. Surveillance Summaries* , *70*(35), 1214. https://doi.org/10.15585/mmwr.mm7035e2

Lan, Y., Desjardins, M. R., Hohl, A., & Delmelle, E. (2021). Geovisualization of COVID-19: State of the art and opportunities. *Cartographica The International Journal for Geographic Information and Geovisualization*, *56*(1), 2–13. https://doi.org/10.3138/cart-2020-0027

Lan, Y., Tang, W., Dye, S., & Delmelle, E. (2020). A web-based spatial decision support system for monitoring the risk of water contamination in private wells. *Annals of GIS*, *26*(3), 293–309. https://doi.org/10.1080/19475683.2020.1798508

Lee, E. K., Pietz, F. H., Chen, C.-H., & Liu, Y. (2017). An interactive web-based decision support system for mass dispensing, emergency preparedness, and biosurveillance. *Proceedings of the 2017 International Conference on Digital Health*, 137–146. https://doi.org/10.1145/3079452.3079473

Liu, P., Ibaraki, M., VanTassell, J., Geith, K., Cavallo, M., Kann, R., Guo, L., & Moe, C. L. (2021). A sensitive, simple, and low-cost method for COVID-19 wastewater surveillance at an institutional level. *The Science of the Total Environment*, 151047.

https://doi.org/10.1016/j.scitotenv.2021.151047

Malczewski, J. (1999). *GIS and Multicriteria Decision Analysis*. Hoboken, NJ: John Wiley & Sons.

Marakas, G. M. (2003). *Decision support systems in the 21st century (Vol. 134)*. Upper Saddle River, NJ: Prentice Hall.

Masrur, A., Yu, M., Luo, W., & Dewan, A. (2020). Space-Time Patterns, Change, and Propagation of COVID-19 Risk Relative to the Intervention Scenarios in Bangladesh. *International Journal of Environmental Research and Public Health*, *17*(16), 5911. https://doi.org/10.3390/ijerph17165911

Medema, G., Been, F., Heijnen, L., & Petterson, S. (2020). Implementation of environmental surveillance for SARS-CoV-2 virus to support public health decisions: Opportunities and challenges. *Current Opinion in Environmental Science & Health*, *17*, 49–71. https://doi.org/10.1016/j.coesh.2020.09.006

Mwaura, D., & Kada, M. (2017). Developing a web-based spatial decision support system for geothermal exploration at the Olkaria geothermal field. *International Journal of Digital Earth*, *10*(11), 1118–1145. https://doi.org/10.1080/17538947.2017.1284909

Naughton, C. C., Roman, F. A., Alvarado, A. G. F., Tariqi, A. Q., Deeming, M. A., Bibby, K., Bivins, A., Rose, J. B., Medema, G., Ahmed, W., & Others. (2021). Show us the data: Global COVID-19 wastewater monitoring efforts, equity, and gaps. *medRxiv*. https://doi.org/10.1101/2021.03.14.21253564

NC government. (2021a, February 24). *Governor Cooper Announces Easing of COVID-19 Restrictions as North Carolina Trends Stabilize*. https://governor.nc.gov/news/press-

releases/2021/02/24/governor-cooper-announces-easing-covid-19-restrictions-north-carolina-trends-stabilize

NC government. (2021b, March 23). *Gov. Cooper Announces North Carolina Will Relax Some COVID-19 Restrictions*. https://governor.nc.gov/news/press-releases/2021/03/23/gov-cooper-announces-north-carolina-will-relax-some-covid-19-restrictions

NSF. (2007). *Cyberinfrastructure Vision for 21st Century Discovery*. National Science Foundation, Cyberinfrastructure Council. https://www.nsf.gov/pubs/2007/nsf0728/

Peccia, J., Zulli, A., Brackney, D. E., Grubaugh, N. D., Kaplan, E. H., Casanovas-Massana, A., Ko, A. I., Malik, A. A., Wang, D., Wang, M., Warren, J. L., Weinberger, D. M., Arnold, W., & Omer, S. B. (2020). Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nature Biotechnology*, *38*(10), 1164–1167. https://doi.org/10.1038/s41587-020-0684-z

Pelekis, N., Theodoulidis, B., Kopanakis, I., & Theodoridis, Y. (2004). Literature review of spatio-temporal database models. *Knowledge Engineering Review*, *19*(3), 235–274. https://doi.org/10.1017/s026988890400013x

Peng, Z.-R., & Tsou, M.-H. (2003). *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. John Wiley & Sons.

Peuquet, D. J., & Duan, N. (1995). An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Systems*, *9*(1), 7–24. https://doi.org/10.1080/02693799508902022

Prado, T., Fumian, T. M., Miagostovich, M. P., & Gaspar, A. M. C. (2012). Monitoring the hepatitis A virus in urban wastewater from Rio de Janeiro, Brazil. *Transactions of the Royal*

Society of Tropical Medicine and Hygiene, *106*(2), 104–109. https://doi.org/10.1016/j.trstmh.2011.10.005

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *26*(1), 43–49. https://doi.org/10.1109/tassp.1978.1163055

Sugumaran, R., & Degroote, J. (2010). *Spatial Decision Support Systems: Principles and Practices*. Boca Raton, FL: CRC Press.

Sweetapple, C., Melville-Shreeve, P., Chen, A. S., Grimsley, J. M., Bunce, J. T., Gaze, W., Fielding, S., & Wade, M. J. (2022). Building knowledge of university campus population dynamics to enhance near-to-source sewage surveillance for SARS-CoV-2 detection. *The Science of the Total Environment*, *806*, 150406. https://doi.org/10.1016/j.scitotenv.2021.150406

Tambini, G., Andrus, J. K., Marques, E., Boshell, J., Pallansch, M., de Quadros, C. A., & Kew, O. (1993). Direct detection of wild poliovirus circulation by stool surveys of healthy children and analysis of community wastewater. *The Journal of Infectious Diseases*, *168*(6), 1510–1514. https://doi.org/10.1093/infdis/168.6.1510

Tang, W., Feng, W., Jia, M., Shi, J., Zuo, H., Stringer, C. E., & Trettin, C. C. (2017). A cyber-enabled spatial decision support system to inventory Mangroves in Mozambique: coupling scientific workflows and cloud computing. *International Journal of Geographical Information Science*. https://doi.org/10.1080/13658816.2016.1245419

Tayyebi, A., Meehan, T. D., Dischler, J., Radloff, G., Ferris, M., & Gratton, C. (2016). SmartScape™: A web-based decision support system for assessing the tradeoffs among

multiple ecosystem services under crop-change scenarios. *Computers and Electronics in Agriculture*, *121*, 108–121. https://doi.org/10.1016/j.compag.2015.12.003

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*(4), 267–276. https://doi.org/10.1007/BF02289263

Xu, F., & Beard, K. (2021). A comparison of prospective space-time scan statistics and spatiotemporal event sequence based clustering for COVID-19 surveillance. *PloS One*, *16*(6), e0252990. https://doi.org/10.1371/journal.pone.0252990

964 **Appendix**

965 Appendix 1. Sources of the information about the University of North Carolina at Charlotte
966 (retrieved year: 2021).

Sources	URLs
Faculty and Staff Resources	https://www.charlotte.edu/gateway/faculty-staff
Housing and Residence Life	https://housing.charlotte.edu/
University Catalogs	https://catalog.uncc.edu/preview_program.php?catoid=30&poid=8179
Housing and Residence Life	https://housing.charlotte.edu/housing-options/find-your-home
Undergraduate Admissions	https://admissions.charlotte.edu/about-unc-charlotte/university-profile

967

968