

# 1 Using geographic information systems to link population estimates to wastewater surveillance data 2 in New York State, USA

## 3 Linking population estimates to wastewater surveillance data

4

5 Dustin T. Hill<sup>1</sup> and David A. Larsen<sup>1</sup>

6 <sup>1</sup>Department of Public Health, Syracuse University, Syracuse NY, 13244

7 *Correspondence to:* Dustin T. Hill ([dthill@syr.edu](mailto:dthill@syr.edu))

8

### 9 **Abstract.**

10 Sewer systems provide many services to communities that have access to them beyond removal of waste and wastewater.  
11 Understanding of these systems' geographic coverage is essential for wastewater-based epidemiology (WBE), which requires  
12 accurate estimates for the population contributing wastewater. Reliable estimates for the boundaries of a sewer service area or  
13 sewershed can be used to link upstream populations to wastewater samples taken at treatment plants or other locations within  
14 a sewer system. These geographic data are usually managed by public utilities, municipal offices, and some government  
15 agencies, however, there are no centralized databases for geographic information on sewer systems. We created a database for  
16 all municipal sewersheds in New York State for the purpose of supporting statewide wastewater surveillance efforts to support  
17 public health. We used a combination of public tax records with sewer access information, physical maps, and municipal  
18 records to organize and draw digital boundaries compatible with geographic information systems. The methods we employed  
19 to create these data will be useful to inform similar efforts in other jurisdictions and the data have many public health  
20 applications as well as being informative for water/environmental research and infrastructure projects.

21 **Keywords:** Public health; environmental surveillance; sewersheds; wastewater treatment plant; digital infrastructure

22

## 23 **1 Introduction**

24 Wastewater from a community provides a wealth of information about that population including types of infectious diseases  
25 circulating, types of drugs consumed, and types of food eaten (1,2). Wastewater-based epidemiology (WBE) is the use of  
26 wastewater to understand public health issues through testing wastewater samples and then estimating community-level trends  
27 (1,2). WBE became exponentially important during the emergence of COVID-19 with governments using wastewater  
28 surveillance to track the SARS-CoV-2 virus to inform disease spread, intensity, and genetic variation (3). Beyond COVID-19,  
29 WBE has a strong record of informing our understanding of antimicrobial resistance (4), expanding the study of endemic  
30 diseases, such as influenza and HIV, and guiding public health research and policy into the future (5). In addition to providing  
31 information on infectious disease threats to public health (6), testing wastewater can inform understanding of the burden of  
32 opioid and other drug use in a community (7). Linking the population to the wastewater sample is a key step in WBE (8,9), as  
33 accurate population estimates are needed to understand what is found in wastewater in terms of population-level burden and  
34 create statistical models that use these data (10). Linking the sample to the population is also necessary for precise public  
35 health intervention if an outbreak is detected (11).

36

37 The area from which a treatment plant receives wastewater is known as the sewershed, defined as an area of land where all the  
38 sewers flow into one point (12). Sewersheds are usually polygon shapes inclusive of the property parcels that have connections  
39 to the sewer system but can also include larger portions of land area that may contribute runoff rainwater to the system (13).  
40 Sewersheds follow natural hydrology (like rivers and streams), flowing downhill congruent with topography in most cases.  
41 There are, however, sewer systems that “jump” rivers with pipes crossing the banks aboveground or going underneath the  
42 waterway. Systems can also include what are known as force mains, which are sewer lines that force water uphill to the final  
43 collection system (14). This variation complicates the estimation of boundaries for sewer systems when survey maps do not  
44 exist, but there are ways to approximate boundaries such as using municipal borders and tax registries that collect information  
45 on whether a parcel of land is connected to public sewer or not (15). The sewershed boundary can also be linked with other  
46 geographically-referenced data, such as population census data, to estimate the population within a sewershed (16), and

47 linkages can be made with environmental data, such as precipitation and temperature, to better understand wastewater test  
48 results. Linking wastewater samples to geographically relevant data can enhance research on infectious disease transmission  
49 and improve the utility of data related to the social determinants of health.

50

51 Geographic data on sewer systems, including manhole locations, sewer lines, and related infrastructure, are usually maintained  
52 by the municipality that owns the system or the engineers that designed it originally (DEC, personal communication). In other  
53 words, centralized databases for sewers across different jurisdictions are uncommon and likely unavailable (NY GIS, personal  
54 communication). As New York State's wastewater surveillance network increased in scale and scope, a centralized database  
55 for all of New York State's municipal sewer systems was needed to link testing of municipal wastewater to upstream  
56 populations for public health surveillance of SARS-CoV-2, other pathogens, and substances of public health interest. Since no  
57 centralized database existed beyond a list of the point locations for each treatment plant per discharge permitting records (17),  
58 creation of a database of sewersheds was necessary geographic data for the state's wastewater surveillance network. The final  
59 product is a spatial database that has every municipal sewer system in New York State mapped and linked to the treatment  
60 plant and population served by that sewer system. In what follows, we detail the methods for gathering and drawing the final  
61 sewershed polygons, and steps for data validation and quality assurance. These methods can be applied in other jurisdictions  
62 with similar needs and we explore some of the uses for these data beyond public health.

63

## 64 **2 Data and Methods**

### 65 **2.1 Data sources**

#### 66 **2.1.1 Treatment plant operators**

67 The NY Department of Environmental Conservation (DEC) provided a database listing 638 municipal wastewater treatment  
68 plants (17). This list included the name of the facility, its location via street address and geocoordinates, as well as other  
69 metadata such as permitted discharge capacity. The list of these plants were the targets for this project with the goal being to  
70 find and draw sewershed boundaries for each plant. Contact information for 385 of the plants on the list was provided by the  
71 DEC, and a survey was distributed to these plants via email and phone in 2021 (see Hill et al., 2022 for the full question list).

72 The purpose of the survey was to determine what data the treatment plant operators had regarding the service area of the facility  
73 and what they knew about WBE methods for tracking substances in wastewater for public health. Results on operator  
74 knowledge regarding WBE has been published previously (18). If no data existed, the operator provided a description of the  
75 service area to facilitate drawing of the boundary using other data sources. The respondents were followed up with email and  
76 phone calls to request they send the data for inclusion in the database. Data were received in different formats including digital  
77 shapefiles of the boundary, sewer manholes, and/or sewer mains; physical maps of the boundary, sewer manholes, and/or  
78 sewer mains; and address lists of the properties connected to the sewer system.

### 79 **2.1.2 New York State Parcel database**

80 NY maintains a geodatabase of property tax data for all properties in the state. While data for several counties is publicly  
81 available, most data are not publicly accessible. We received permission from the state Geographic Information Systems office  
82 for use of the data to aid in construction of the sewershed boundaries. The tax parcel data consisted of spatial polygons and  
83 included a field indicating if the parcel was on public sewer or a private system. This allowed classification of parcels based  
84 on sewer type. In addition, special district tables were provided that included information on water and sewer districts that  
85 could be joined to the tax parcels by parcel ID number. The sewer district data was not complete for all counties, but, for  
86 counties where it was available, it was used to classify separate sewer systems that were adjacent. The version of the data used  
87 was from 2020 (19).

### 88 **2.1.3 Municipal websites and other data sources**

89 In addition to calling municipalities, we researched public sewer systems online through town, village, and city websites  
90 searching for any published records, reports, or information on the location of sewer service areas. From public websites, we  
91 were able to obtain copies of physical maps and descriptions of some sewer systems. A list of websites for source material is  
92 provided in the supplementary material.

93

94 For the three counties in northern New York, additional data were provided by the Development Authority of the North  
95 Country (DANC), which is a public benefit corporation set up to manage the infrastructure needs for those counties (20).

96 DANC provided estimates for 55 sewershed boundaries based on work they completed using village boundaries, sewer billing  
 97 data, and sewer district data. Further description is available in the supplemental material.

98 **2.2 Combination of data**

99 Data were collected and compiled for each county in NY using eight methods listed in Table 1. Wastewater treatment plants  
 100 (WWTPs) were mapped to the county to provide the initial locations for communities that would have access to public sewer.  
 101 The first sewersheds that were added were those that were provided by the treatment plant or municipal/county officials and  
 102 these were not modified. Following these, if we had an address list for properties that were billed for public sewer by the  
 103 municipality, we matched parcels from the NY tax parcel data for the county. The boundaries for the parcels were then  
 104 dissolved to form the final boundary.

105  
 106  
 107

**Table 1: Methods for digitizing sewersheds**

<i>Method</i>	<i>Method description</i>	<i>Count</i>	<i>Percent</i>	<i>Median flow (mgd)</i>	<i>Mean flow (mgd)</i>
<b>Digitized from address list</b>	Digital shapefile boundary for the sewershed was provided by a third party (i.e., the treatment plant operator or owning municipality).	1	0.2	0.09	0.09
<b>Digitized from manhole/sewer main shapefile</b>	NY State Tax Parcel database was used to estimate parcels served by this sewer system. Parcels that are cross listed on the address list provided were assigned to the sewershed to draw the boundary.	35	5.8	0.40	1.93
<b>Digitized from physical map (JPEG/PDF)</b>	Manhole and or sewer main data were sent by the treatment plant operator and the boundary for the service area was created using these data and NY State Parcel data that intersected with the manhole/sewer main data.	126	21.0	0.52	3.77
<b>Parcel digitized from DANC records</b>	A document containing a map of the region served by the treatment plant was used to draw the boundary using ArcGIS software.	55	9.2	0.10	1.26
<b>Parcel digitized from description</b>	Development Authority of the North County (DANC) provided a shapefile of sewersheds that were created starting in 2010 and updated annually. The version this project used was obtained in 2021. Boundary matches current estimates and therefore DANC boundary was used for this treatment plant’s sewershed.	19	3.2	0.99	3.30

<b>Parcel digitized no description</b>	Treatment plant operator provided a text description of the service area including roads, natural boundaries, number of influent points, and towns/villages/cities served. The boundary was drawn using ArcGIS software and Census data for major roads, municipal boundaries, and NY State Tax Parcel database. Tax parcels that are recorded as paying a sewer tax and fall within the described service area were assigned to this sewershed.	304	50.8	0.35	1.34
<b>Provided by treatment plant/municipality/county</b>	New York State Tax Parcel database was used to estimate the parcels served by this sewer system. Parcels that are recorded as paying a sewer tax and proximate to the wastewater treatment plant's geolocation were assigned to the sewershed. Adjacent sewer systems were subdivided using sewer district data when available.	58	9.7	11.00	47.14
<b>Village boundary</b>	No adjacent parcels were indicated to be on sewer. Municipal village boundary was used as a placeholder.	1	0.2	0.05	0.05
<b>Total number of sewersheds</b>		599 <sup>1</sup> (596 WWTPs)			

*Note: Of the original 638 treatment plants identified in DEC's database, sewersheds were drawn for 592 of them with 4 additional identified that were not in the database. There were 46 that we did not draw sewersheds for because they either do not exist anymore or are included in another system.*

108

109

110 For municipalities that we received digital manhole or sewer main data, we mapped those against the tax parcel data for the

111 county and selected the parcels that were adjacent to sewer mains and proximate to manhole locations (Fig. 1a and Fig. 1b).

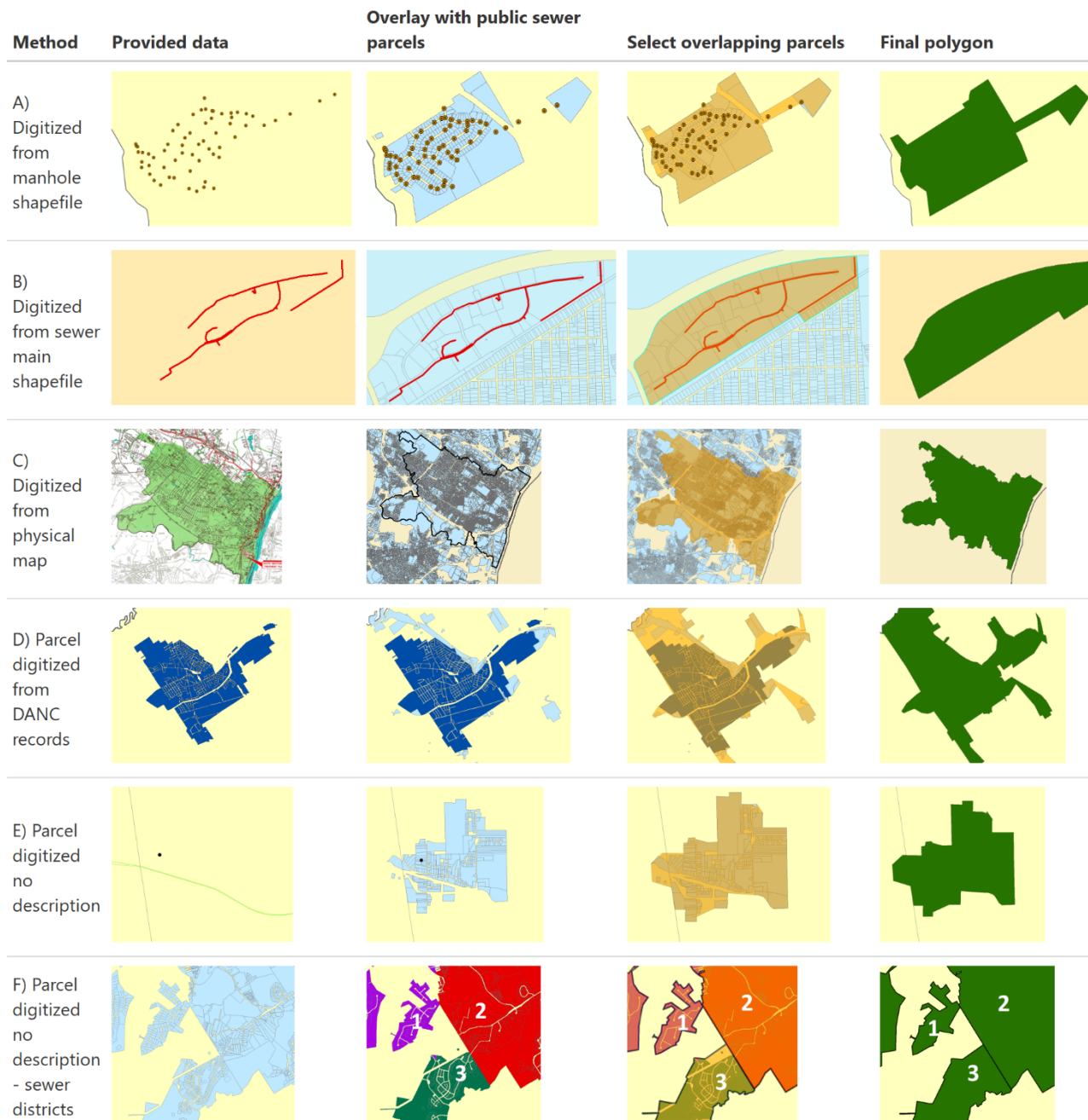
112 These parcels were then dissolved to form the sewershed boundary and linked to the municipality's treatment plant. The other

113 type of data provided were physical maps showing the boundaries of the service area. For these maps, we aligned the borders

114 of the map with the tax parcel data for that county and municipality and then selected the parcels within the boundary (Fig.

115 1c). Listed roads and other landmarks that formed the boundary were used to ensure the correct parcels were selected. The

116 parcel boundaries were then dissolved to form the final sewershed boundary.



117

118 **Figure 1: Methods used to create sewersheds.**

119

120 Boundaries provided by DANC were compared to the reported parcels on public sewer. For many municipalities, the  
121 boundaries corresponded well and the DANC boundary was used, however, in some cases adjacent parcels were reported to  
122 be on public sewer but not in the DANC sewershed. Therefore, we added these additional parcels to form the final sewershed  
123 combining the DANC sewershed with adjacent parcels on public sewer for the same municipality (Fig. 1d).

124

125 For the remaining municipalities that we did not have data provided or could not find public records for sewer boundaries or  
126 districts, we drew the boundaries based on the parcels on public sewer near the geolocation of the municipal WWTP (Fig. 1e).

127 For many locations, the clusters of sewer parcels corresponded well to the locations of treatment plants and there were clear  
128 boundaries between sewer systems. For some counties with several WWTPs serving adjacent communities, it was harder to

129 distinguish between the sewer parcels and the correct WWTP to link it to. For those locations, we examined the sewer district  
130 data from the tax parcel special district tables. Where available, sewer districts were identified and grouped together to form

131 distinct sewer systems that were then linked to the correct WWTP (Fig. 1f). Some counties did not have good sewer district  
132 data and for those the treatment plant or municipal official was contacted to request information on the boundaries for adjacent

133 sewer systems.

### 134 **2.3 Adding population data**

135 Estimates for 2020 population by sewershed were calculated using R statistical software version 4.1.1 (21). We first estimated  
136 the 2010 population for each sewershed using an overlay of 2010 U.S. Census blocks on top of the sewershed boundaries.

137 Census blocks were selected because they represent the smallest geometry for public census data and increasing the accuracy  
138 for the intersection with sewershed boundaries. We calculated the proportion of the area for partial block overlap and then

139 assigned the proportional 2010 decennial population of the block to the sewershed assuming equal distribution of the  
140 population in the blocks. The apportioned values were then aggregated to obtain a total population estimate for the sewershed.

141

142 Then, we repeated this procedure using 2010 decennial population data for the block group and 2018 American Community  
143 Survey (ACS) data for the block group to get 2010 and 2018 population by sewershed based on block groups. We used these

144 values to estimate the annual rate of population change per sewershed using the following formula:



145

$$\text{Annual growth rate} = \left( \frac{2018 \text{ pop} - 2010 \text{ pop}}{2010 \text{ pop}} \right) / 8 \text{ years} \quad (1)$$

146

147 We then applied this average annual change to the sewershed population based on the block data from 2010 and estimated the  
148 population after ten years of growth using the following formula to calculate 2020 population estimates:

149

$$2020 \text{ pop est} = 2010 \text{ pop} * (1 + \text{annual growth rate})^{10} \quad (2)$$

150

151 We obtained all population estimates using the “tidycensus” R package (22) and we use the “tigris” package (23) for U.S.  
152 Census geometry data. R code for US census intersections is publicly available at [https://github.com/dthill196/NY-Sewershed-](https://github.com/dthill196/NY-Sewershed-Populations)  
153 [Populations](https://github.com/dthill196/NY-Sewershed-Populations).

#### 154 **2.4 Validation and quality assurance**

155 We validated sewershed boundaries by cross referencing data types when multiple forms were available such as having  
156 physical maps and shapefiles. For sewersheds where the boundaries were in question or more difficult to estimate, direct  
157 contact and consultation with treatment plant operators was used to check boundaries against the knowledge of the treatment  
158 plant staff. Short of formal land surveys of the municipality, the data we created are not easily validated. To check accuracy  
159 for all sewersheds, we ran correlations of the population estimates, population density, and surface area of the final polygons  
160 with the permitted discharge capacity. Discharge capacity increases with population served and was one way to check that  
161 methods for creating sewersheds were producing good estimates for areas served, and the method has been used by other  
162 researchers working on population and treatment plant linkages (24).

#### 163 **3 Results**

164 We received 116 completed surveys with 19 responses that had a description of the service area. These were then combined  
165 with other data sources, such as the NY Tax Parcel database, to draw the sewershed boundary. The remaining 97 responses  
166 indicated that the plant had data on the service area of the plant. The survey respondent was then contacted requesting they

167 share the data with our research team. We received digital data for the boundaries of 38 treatment plants and digital data for  
168 manholes and sewer mains for 35 treatment plants, and an address list for one treatment plant. Data were retrieved in a  
169 combination of survey call-backs and contact with treatment plants after the survey ended (Table 1). In total, 294 sewersheds  
170 were digitized using data provided or descriptions of the system representing 49.1 percent of the final database for NY  
171 sewersheds and these facilities were among the largest in the state with a median flow rate of 11 millions of gallons per day  
172 (mgd) and mean of 47 mgd (Table 1).

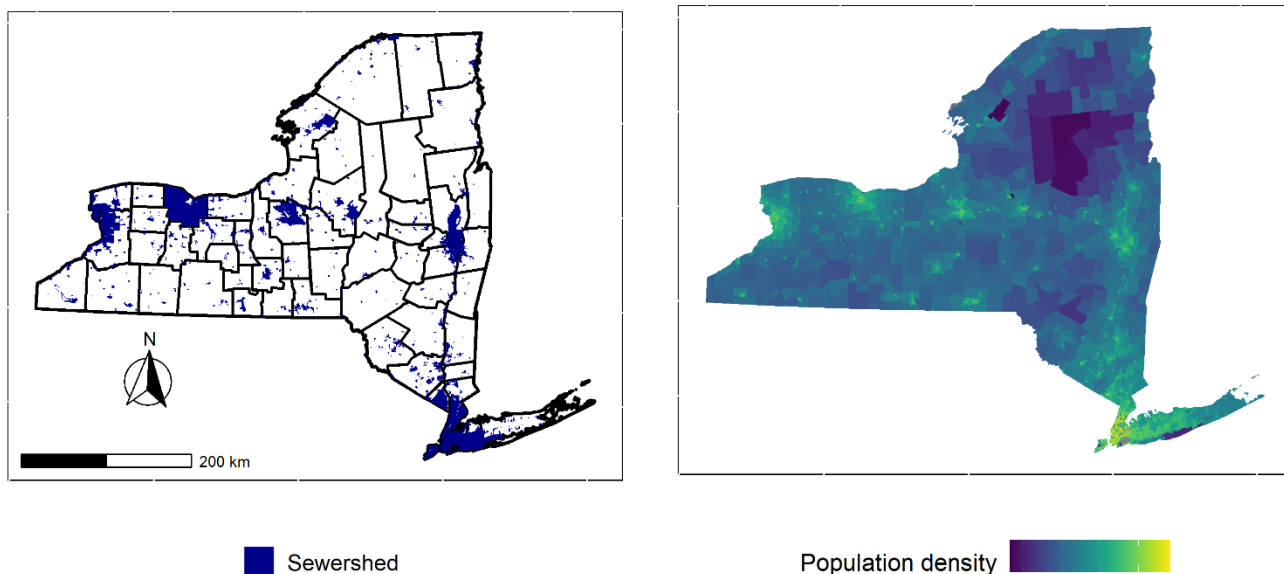
173

174 The remaining sewersheds (304) were digitized from NY tax parcel data and sewer district data with one sewershed digitized  
175 using the village municipal boundary (Table 1). Of the original 638 municipal treatment plants identified using the DEC  
176 database, we were able to create boundaries for 592 of them. There were 46 WWTPs in the DEC list that we did not create  
177 sewershed boundaries for because they either did not exist anymore or were additional permits for coincident treatment plants  
178 representing combined sewer overflow facilities or large pump stations (see supplemental material for a list of permits and  
179 reasons they were excluded). In addition, we created sewershed boundaries for 4 additional WWTPs not in the DEC database.  
180 Two of these facilities are owned by Native American communities and are on land owned by those communities with their  
181 permits supervised by Federal agencies. We included them because their boundaries were evident from tax parcel data for  
182 public sewer that were not linked to any treatment plant in the DEC list. Two additional facilities were reported by county  
183 governments with their permit information and physical maps provided showing their locations. The final count for our  
184 database is 596 municipal WWTP sewersheds. Three WWTPs have two sewersheds each representing separate influent  
185 streams from different parts of their service area bringing the total number of sewersheds to 599 (Table 1).

186

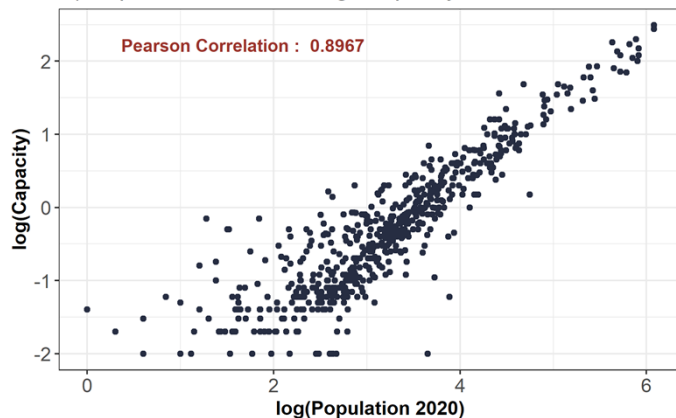
187 The treatment plants are distributed around the state with some regions and counties having many separate sewer systems with  
188 others having fewer (Fig. 2). The total population in the state estimated to be on public sewer is over 16 million, which is  
189 approximately 82.6 percent of the total state population (using 19.5 million as the estimated NY population in 2020).  
190 Population coverage per sewershed ranged between less than 100 people to as high as 1.2 million. Population density ranged  
191 between 1.6 people per km<sup>2</sup> and 28,000 people per km<sup>2</sup>. Surface area ranged between 58,000 m<sup>2</sup> and 990 km<sup>2</sup>. There was high

192 correlation between the permitted discharge capacity of each facility and estimated population served (Pearson correlation =  
193 0.897, Fig. 3), moderately high correlation between discharge capacity and estimated population density (Pearson correlation  
194 = 0.569, Fig. 3), and high correlation between discharge capacity and surface area (Pearson correlation =0.852, Fig. 3). When  
195 examining correlations across the different methods for creating sewersheds (Table 2), we find high all Pearson correlation  
196 values higher than 0.5 and statistically significant (p-value < 0.05) except for one method. The use of manhole and sewer main  
197 data was correlated significantly with population and surface area, but for population density, we found lower correlation with  
198 a Pearson estimate of 0.195 and p-value = 0.27 (Table 2).

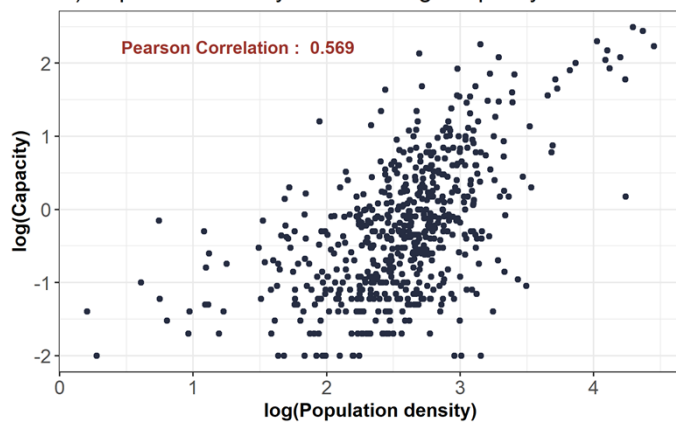


It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

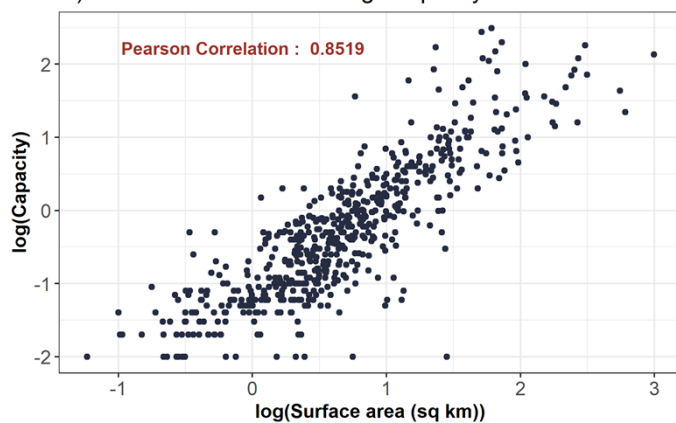
A) Population and discharge capacity



B) Population density and discharge capacity



C) Surface area and discharge capacity



202

203 **Figure 3: Correlations between permitted discharge capacity of WWTPs and A) estimated population, B) population density, and**

204 **C) surface area of the sewershed.**

205

**Table 2: Correlation results for each method**

<i>Method</i>	<i>Discharge capacity and estimated population</i>		<i>Discharge capacity and population density</i>		<i>Discharge capacity and surface area</i>	
	<i>Pearson correlation</i>	<i>p-value</i>	<i>Pearson correlation</i>	<i>p-value</i>	<i>Pearson correlation</i>	<i>p-value</i>
<b>Provided by treatment plant/municipality/county</b>	0.966	<0.01	0.675	<0.01	0.673	<0.01
<b>Parcel digitized no description</b>	0.83	<0.01	0.439	<0.01	0.826	<0.01
<b>Parcel digitized from description</b>	0.974	<0.01	0.615	<0.01	0.853	<0.01
<b>Digitized from physical map (JPEG/PDF)</b>	0.876	<0.01	0.515	<0.01	0.854	<0.01
<b>Digitized from manhole/sewer main shapefile</b>	0.707	<0.01	0.195	0.27	0.668	<0.01
<b>Parcel digitized from DANC records</b>	0.907	<0.01	0.444	0.001	0.842	<0.01

206

#### 207 4 Data Availability

208 Data described in this manuscript is available upon request to the author. Links to publicly available databases used in this

209 project are provided in the reference section and supplemental documents. R code for creating US census intersections is

210 publicly accessible at <https://github.com/dthill196/NY-Sewershed-Populations> .

## 211 **5 Discussion and Conclusions**

### 212 **5.1 Evaluation of sewershed boundaries**

213 Correlation results between the final sewershed polygons, the discharge capacity of the WWTPs, and the three metrics of  
214 population, population density, and surface area were high for all methods used (Figure 3) suggesting that each method  
215 contributed acceptable results. One exception to this were sewersheds drawn using manhole or sewer main shapefile data  
216 (Table 2). This method resulted in the lowest correlation between discharge capacity and population density, however, there  
217 was high correlation for estimated population and surface area for this method. Therefore, despite the one low value for this  
218 method, we feel that the resulting sewershed boundaries are acceptable estimates for the service areas of the treatment plants.  
219 There were 35 sewersheds drawn using this method with the majority (n = 21) from one county. Removing the county in  
220 question from the correlation test yielded a new correlation coefficient of 0.671 (p-value = 0.012) suggesting there might be  
221 greater uncertainty with the sewershed boundaries from that county.

222

223 The county in question is unique in being highly populated but most of the county population not living on public sewer and,  
224 instead of the more typical private septic system for residents, most of the public wastewater goes into what are known as  
225 cesspools (25). The county in question has a population over 1 million people but our estimates report only 397,000 of the  
226 residents are on public sewer meaning it has high population density but low public sewer population, which might lead to the  
227 lower correlation between population density and discharge capacity of the sewer systems. In addition, the county has 30  
228 WWTPs with 22 plants with discharge capacities below 1 million gallons per day. Therefore, the county might be an anomaly  
229 among counties in NY with regards to public sewer systems. Despite the low correlation for population density, the county  
230 has high correlation for population estimates and surface area with discharge capacity suggesting the resulting boundaries are  
231 comparable in their estimates to the rest of the state.

232

233 The resulting database of sewersheds includes 4 treatment plants not in the DEC database. These additional plants were drawn  
234 based on information from the municipalities and because there were public sewer parcels that matched the description by the  
235 municipal officials. DEC's public database could be behind in its updates between discharge permit documents and adding

236 locations with the most recent evaluation of sewer infrastructure from 2008 (26). We also did not create boundaries for 46  
237 treatment plants listed in DEC's database for several reasons including that they do not exist anymore or they are included in  
238 the sewershed of another facility because they are a smaller pump station or overflow facility. A complete list of the facilities  
239 not included and explanations why are provided in the supplemental material.

## 240 **5.2 Applications for public health, research, and policy**

241 The NY Sewershed database was created to support public health efforts related to WBE including the important role of  
242 providing population estimates for wastewater test results (10,16). Linking wastewater results to the upstream population is  
243 essential for interpretation of findings (6) particularly for case and incidence data for health outcomes of concern such as  
244 COVID-19 (27), Polio (28) and other enteroviruses, and opioid use (29). Community-based research that uses wastewater data  
245 needs accurate estimates for population served by the facility and large-scale investigations that need to scale up quickly may  
246 not have the time to create boundaries or provide estimates in a timely manner particularly if the data are to be used for public  
247 health policy. Linking wastewater surveillance data to communities is necessary for policy applications of WBE with  
248 increasing application in the US since the start of the COVID-19 pandemic (3,30). Beyond WBE, this database could be used  
249 to support water quality research, hydrology research, and policy related to infrastructure improvements.

250

251 Water quality researchers might find utility in these data examining freshwater systems and discharge locations. With the  
252 sewershed boundaries, researchers now have estimates for where sewer systems border or even cross waterways enabling  
253 inclusion of these data in research plans as has been done before (9). In addition, hydrology research can use these data to  
254 understand the flow of wastewater to treatment plants from communities and potential relationships with topography and  
255 hydrography (31). Lastly, the infrastructure needed to build and maintain sewer systems is aging in NY according the DEC  
256 (26) with 36.2 billion dollars estimated to repair, replace, or update sewer systems. Comprehensive mapping of the served  
257 communities through the NY sewershed database might be helpful in state-wide and county-wide planning for consolidating  
258 sewer systems. While not an official survey of sewer systems, the database we created provides an important  
259 step in identifying current sewer systems in NY.

### 260 **5.3 Limitations and future directions**

261 The sewersheds that we created come with some limitations. For all data created, there is uncertainty around the estimated  
262 boundaries because they were not produced from formal surveys, but instead methods that estimate the boundary based on  
263 secondary data sources. Physical maps that were used might be the most accurate followed by digital data provided by  
264 treatment plants, however, the date of some of the maps' original drawing go back several years, sometimes to the original  
265 building of the treatment plant. In addition, the tax parcel data are up-to-date but come with limitations of their own including  
266 disclaimers regarding their accuracy. Therefore, these data should not be used in place of official surveys or assessments of  
267 sewer systems for the purpose of projects related to infrastructure updates. Instead, these data can support where surveys could  
268 be done and help locate sewer systems if previously unmapped.

269

270 Further, US Census data estimates represent the population at one time point and data may not be the best match for some  
271 locations that are tourist destinations or have frequently population changes due to seasonal residents. Locations of this kind  
272 should use the provided population estimates adjusted for their research endeavor and where necessary seek better estimates  
273 from other sources. In addition, low population estimates were obtained for many locations (56 sewersheds had population  
274 estimates below 100). These low population levels are due to the limits built into the US Census data where population numbers  
275 may be suppressed due to low number of residents in rural communities. Similar to sewersheds with population fluctuation,  
276 alternate sources may be needed to improve estimation for small sewersheds such as seeking information directly from  
277 municipalities on number of sewer hookups.

278

279 The future of these data will require updates as time progresses with the potential for sewer systems to be consolidated and  
280 new systems to be built making current estimates less reliable. Updates can be done using revised tax parcel data showing  
281 additional parcels on sewer as well as surveys of treatment plant operators and municipalities to identify any changes. Methods  
282 described in this paper can be applied in other jurisdictions including surveying treatment plants for data availability and  
283 working with public and private repositories of land use and property data to estimate Sewershed boundaries. In addition, these  
284 methods can be used to aid in formulation of upstream and community-level sampling approaches for public health where



285 sewersheds are sub-divided into smaller units to better identify outbreaks. The potential for changes in sewer infrastructure is  
286 a potential challenge, but the current dataset and methods are valuable in their comprehensiveness, novelty, and utility to  
287 support public health, water research, and policy.

## 288 **Acknowledgements**

289 This project was made possible by the CDC's Environmental Public Health and Emergency Response Program, NYS Unique  
290 Federal Award Number NUE1EH001341 (NYS Environmental Public Health Tracking Network Maintenance and  
291 Enhancement to Accommodate Sub-County Indicators). We would also like to thank Ed Hampston, AJ Smith, and Eric  
292 Weigert for their contributions regarding WWTP locations and contact information. Also, a thank-you to Hannah Cousins,  
293 Bryan Dandaraw, Catherine Faruolo, Alex Godinez, Sythong Run, Simon Smith, Megan Willkens, and Shruti Zirath for their  
294 help calling treatment plants. Additional thanks to Kate Kiyanitsu from the NY GIS office for providing access and support  
295 for NY tax parcel data. Thank-you to Tabassum Insaf, Abigail Stamm, and Jeff Bryant from NY DOH for helping refine some  
296 Sewershed boundaries. Finally, thank-you to Star Carter at the Development Authority of the North Country for providing  
297 estimates for municipal sewer districts for several treatment plants.

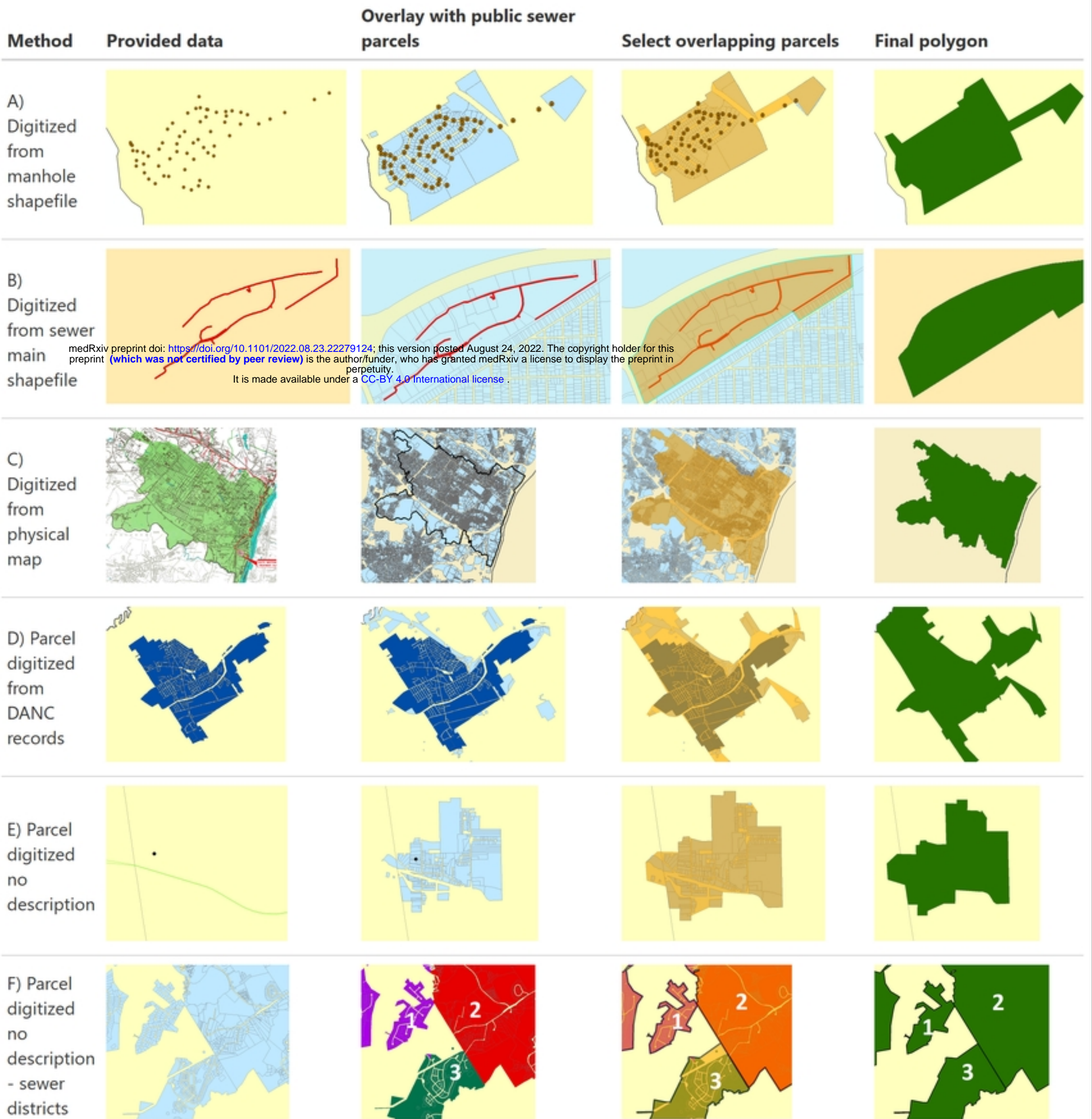
## 298 **References**

- 299 1. Lorenzo M, Picó Y. Wastewater-based epidemiology: current status and future prospects. *Curr Opin Environ Sci Health*.  
300 2019 Jun 1;9:77–84.
- 301 2. Sims N, Kasprzyk-Hordern B. Future perspectives of wastewater-based epidemiology: Monitoring infectious disease  
302 spread and resistance to the community level. *Environ Int*. 2020 Jun 1;139:105689.
- 303 3. Kirby AE, Walters MS, Jennings WC, Fugitt R, LaCross N, Mattioli M, et al. Using Wastewater Surveillance Data to  
304 Support the COVID-19 Response — United States, 2020–2021. *Morb Mortal Wkly Rep*. 2021 Sep 10;70(36):1242–4.
- 305 4. Chau K, Barker L, Budgell E, Vihta K, Sims N, Kasprzyk-Hordern B, et al. Systematic Review of Wastewater  
306 Surveillance of Antimicrobial Resistance in Human Populations. 2021 Jun 21 [cited 2021 Jul 26]; Available from:  
307 <https://www.preprints.org/manuscript/202010.0267/v2>
- 308 5. Kilaru P, Hill D, Anderson K, Collins MB, Green H, Kmush BL, et al. Wastewater surveillance for infectious disease: a  
309 systematic review. *Am J Epidemiol*. In press;2021.07.26.21261155.
- 310 6. Melvin RG, Hendrickson EN, Chaudhry N, Georgewill O, Freese R, Schacker TW, et al. A novel wastewater-based  
311 epidemiology indexing method predicts SARS-CoV-2 disease prevalence across treatment facilities in metropolitan and  
312 regional populations. *Sci Rep*. 2021 Nov 1;11(1):21368.
- 313 7. Boogaerts T, Ahmed F, Choi PhilM, Tschärke B, O'Brien J, De Loof H, et al. Current and future perspectives for  
314 wastewater-based epidemiology as a monitoring tool for pharmaceutical use. *Sci Total Environ*. 2021 Oct 1;789:148047.

- 315 8. Li Z, Li X, Porter D, Zhang J, Jiang Y, Olatosi B, et al. Monitoring the Spatial Spread of COVID-19 and Effectiveness  
316 of Control Measures Through Human Movement Data: Proposal for a Predictive Model Using Big Data Analytics. *JMIR*  
317 *Res Protoc.* 2020 Dec 18;9(12):e24432.
- 318 9. Nelson JR, Lu A, Maestre JP, Palmer EJ, Jarma D, Kinney KA, et al. Space-time analysis of COVID-19 cases and SARS-  
319 CoV-2 wastewater loading: A geodemographic perspective. *Spat Spatio-Temporal Epidemiol.* 2022 Aug 1;42:100521.
- 320 10. Hoar C, Li Y, Silverman AI. Assessment of Commonly Measured Wastewater Parameters to Estimate Sewershed  
321 Populations for Use in Wastewater-Based Epidemiology: Insights into Population Dynamics in New York City during  
322 the COVID-19 Pandemic. *ACS EST Water* [Internet]. 2022 Apr 20 [cited 2022 Jun 28]; Available from:  
323 <https://doi.org/10.1021/acsestwater.2c00052>
- 324 11. Xagorarakis I, O'Brien E. Wastewater-Based Epidemiology for Early Detection of Viral Outbreaks. In: O'Bannon DJ,  
325 editor. *Women in Water Quality: Investigations by Prominent Female Engineers* [Internet]. Cham: Springer International  
326 Publishing; 2020 [cited 2022 Jun 28]. p. 75–97. (Women in Engineering and Science). Available from:  
327 [https://doi.org/10.1007/978-3-030-17819-2\\_5](https://doi.org/10.1007/978-3-030-17819-2_5)
- 328 12. Basaldua J. Watersheds and Sewersheds [Internet]. 2020 [cited 2022 Jun 28]. Available from:  
329 <https://urbanecologycenter.org/item/1386-watersheds-and-sewersheds.html>
- 330 13. Sun N, Hong B, Hall M. Assessment of the SWMM model uncertainties within the generalized likelihood uncertainty  
331 estimation (GLUE) framework for a high-resolution urban sewershed. *Hydrol Process.* 2014;28(6):3018–34.
- 332 14. Selvakumar A, Tafuri AN, Morrison R, Sterling R. State of technology for renewal of sewer force mains. *Urban Water J.*  
333 2011 Sep 1;8(5):279–92.
- 334 15. Manson SM, Sander HA, Ghosh D, Oakes JM, Orfield MW Jr, Craig WJ, et al. Parcel Data for Research and Policy.  
335 *Geogr Compass.* 2009;3(2):698–726.
- 336 16. Tscharke BJ, O'Brien JW, Ort C, Grant S, Gerber C, Bade R, et al. Harnessing the Power of the Census: Characterizing  
337 Wastewater Treatment Plant Catchment Populations for Wastewater-Based Epidemiology. *Environ Sci Technol.* 2019  
338 Sep 3;53(17):10303–11.
- 339 17. DEC. Wastewater Treatment Plants | State of New York [Internet]. 2022 [cited 2022 Jun 28]. Available from:  
340 <https://data.ny.gov/Energy-Environment/Wastewater-Treatment-Plants/2v6p-juki>
- 341 18. Hill DT, Cousins H, Dandaraw B, Faruolo C, Godinez A, Run S, et al. Wastewater treatment plant operators report high  
342 capacity to support wastewater surveillance for COVID-19 across New York State, USA. *Sci Total Environ.* 2022 Sep  
343 1;837:155664.
- 344 19. NY GIS. NYS GIS - Parcels [Internet]. 2022 [cited 2022 Jun 28]. Available from: <http://gis.ny.gov/parcels/>
- 345 20. DANC. Danc - Development Authority of the North Country - About [Internet]. 2022 [cited 2022 Jun 28]. Available  
346 from: <https://www.danc.org/about>
- 347 21. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna, Austria: R Foundation for  
348 Statistical Computing; 2021. Available from: <https://www.R-project.org/>
- 349 22. Kyle Walker, Herman M. tidyensus: Load US Census Boundary and Attribute Data as “tidyverse” and ‘sf’-Ready Data  
350 Frames [Internet]. 2022. Available from: <https://CRAN.R-project.org/package=tidyensus>

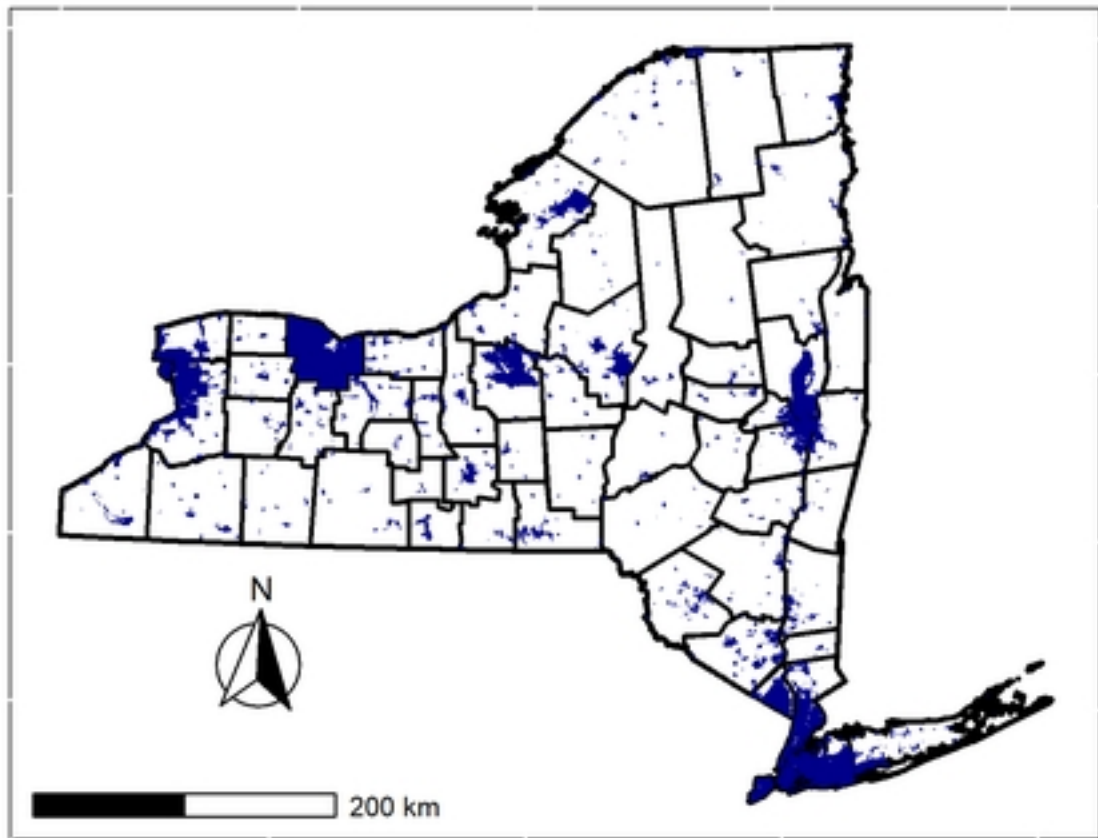
- 351 23. Walker K. tigris: Load Census TIGER/Line Shapefiles [Internet]. 2022. Available from: [https://CRAN.R-](https://CRAN.R-project.org/package=tigris)  
352 [project.org/package=tigris](https://CRAN.R-project.org/package=tigris)
- 353 24. Ehalt Macedo H, Lehner B, Nicell J, Grill G, Li J, Limtong A, et al. Distribution and characteristics of wastewater  
354 treatment plants within the global river network. *Earth Syst Sci Data*. 2022 Feb 9;14(2):559–77.
- 355 25. DEC. Coastal Resiliency and Water Quality in Nassau and Suffolk Counties. 2021;24.
- 356 26. DEC. Wastewater Infrastructure Needs of New York State Report - NYS Dept. of Environmental Conservation [Internet].  
357 2008 [cited 2022 Jun 28]. Available from: <https://www.dec.ny.gov/chemical/42383.html>
- 358 27. Nagarkar M, Keely SP, Jahne M, Wheaton E, Hart C, Smith B, et al. SARS-CoV-2 monitoring at three sewersheds of  
359 different scales and complexity demonstrates distinctive relationships between wastewater measurements and COVID-19  
360 case data. *Sci Total Environ*. 2021 Nov 13;151534.
- 361 28. Link-Gelles R. Public Health Response to a Case of Paralytic Poliomyelitis in an Unvaccinated Person and Detection of  
362 Poliovirus in Wastewater — New York, June–August 2022. *MMWR Morb Mortal Wkly Rep* [Internet]. 2022 [cited 2022  
363 Aug 19];71. Available from: <https://www.cdc.gov/mmwr/volumes/71/wr/mm7133e2.htm>
- 364 29. Wang S, C. Green H, L. Wilder M, Du Q, L. Kmush B, B. Collins M, et al. High-throughput wastewater analysis for  
365 substance use assessment in central New York during the COVID-19 pandemic. *Environ Sci Process Impacts*.  
366 2020;22(11):2147–61.
- 367 30. Reeves K, Liebig J, Feula A, Saldi T, Lasda E, Johnson W, et al. High-resolution within-sewer SARS-CoV-2 surveillance  
368 facilitates informed intervention. *Water Res*. 2021 Oct 1;204:117613.
- 369 31. Shuster WD, Bonta J, Thurston H, Warnemuende E, Smith DR. Impacts of impervious surface on watershed hydrology:  
370 A review. *Urban Water J*. 2005 Dec 1;2(4):263–75.

371

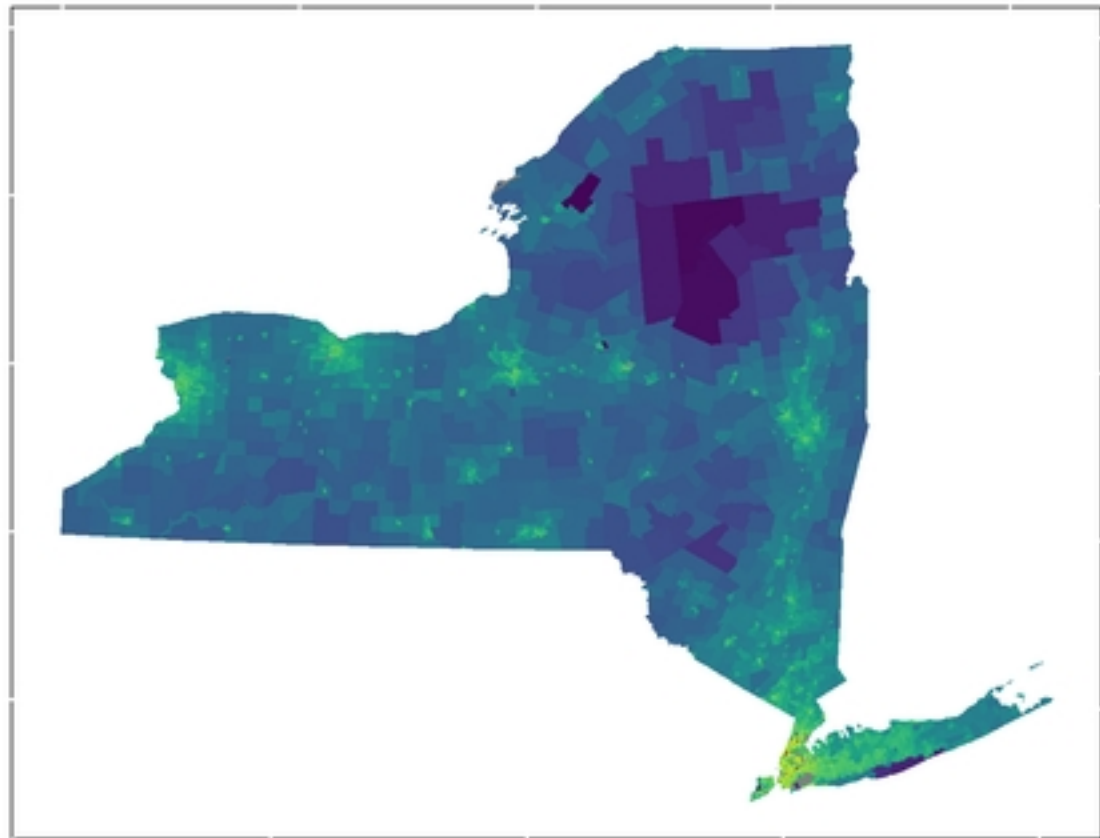


medRxiv preprint doi: <https://doi.org/10.1101/2022.08.23.22279124>; this version posted August 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Figure 1



■ Sewershed




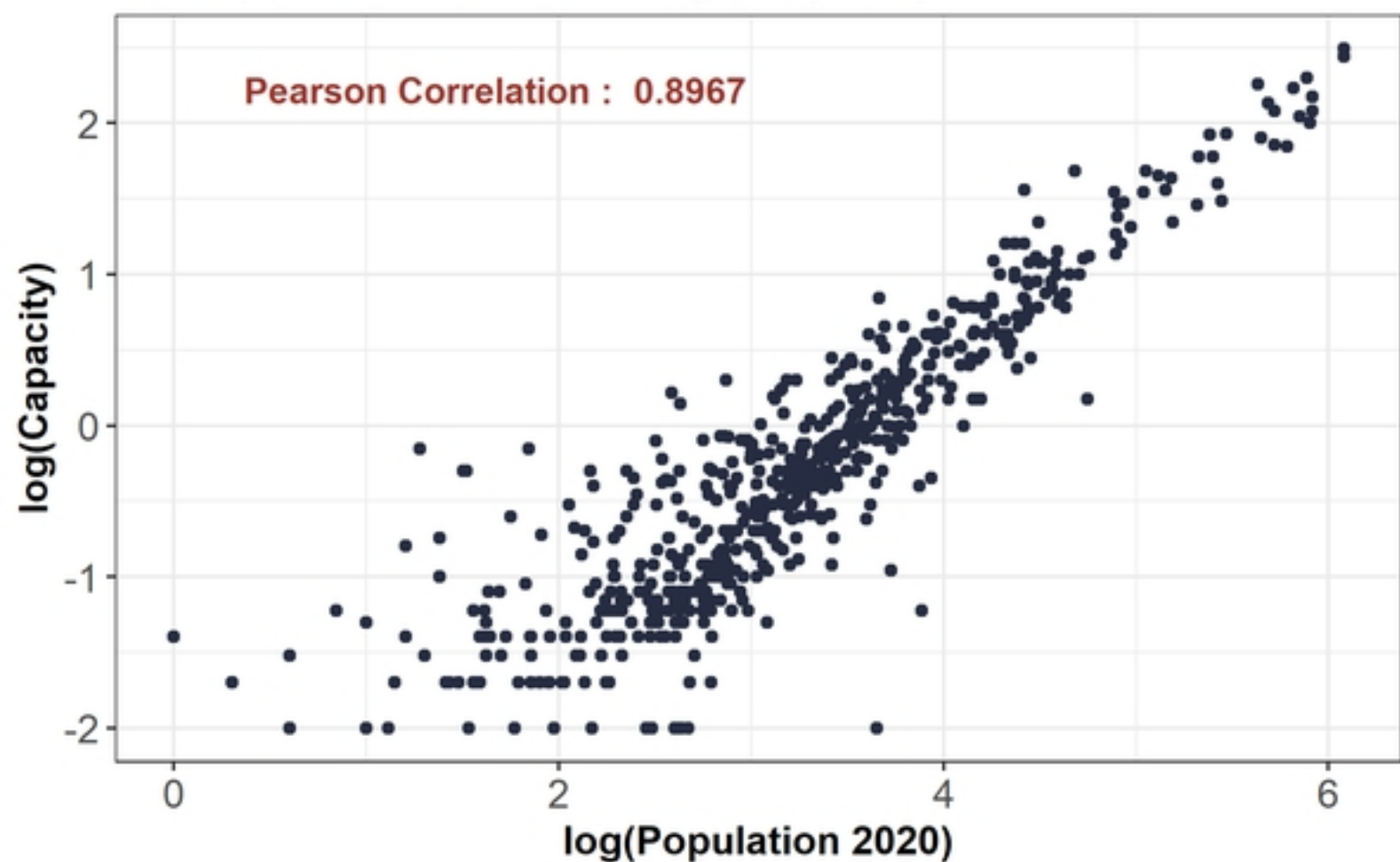
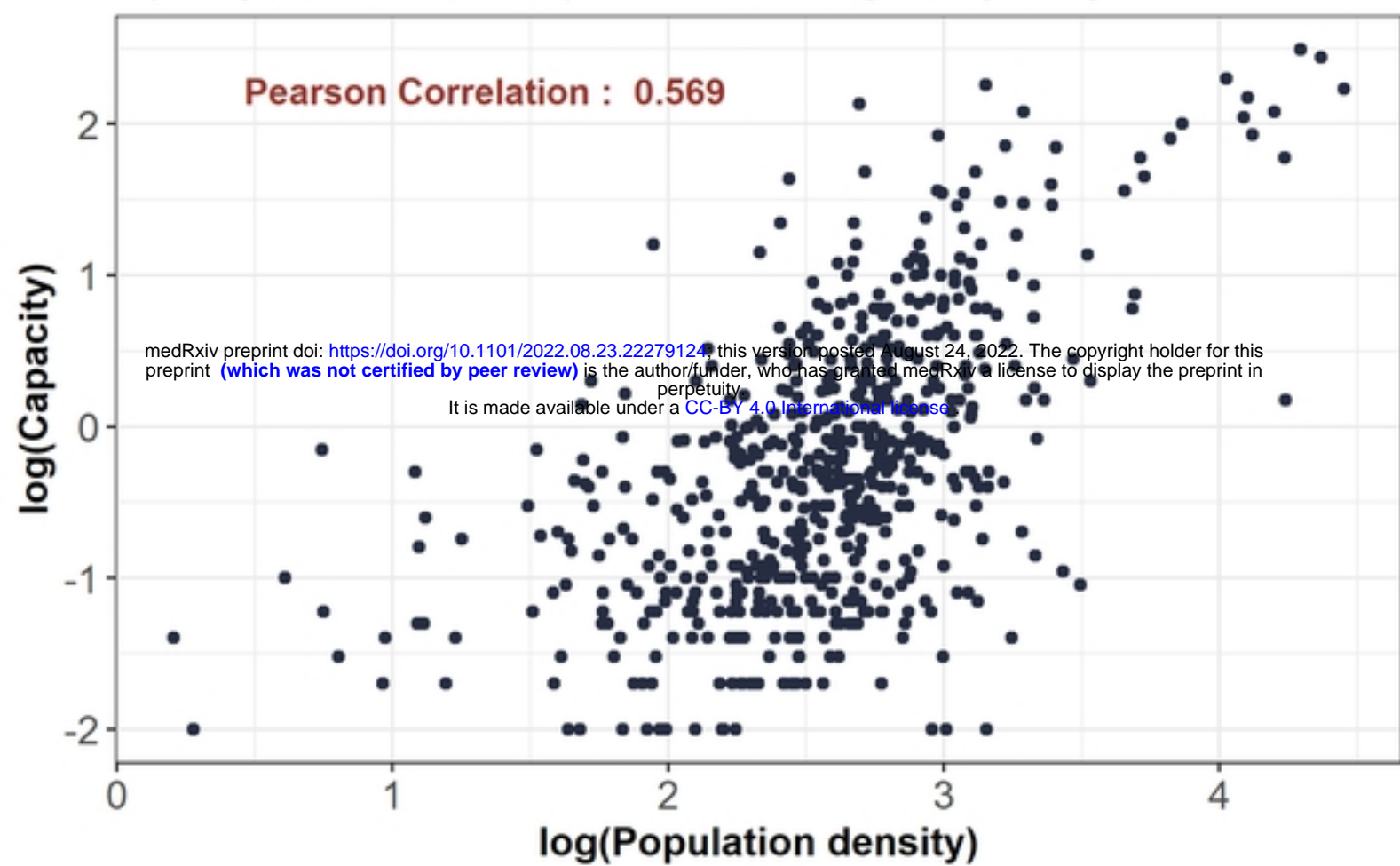
Population density 

Figure 2

### A) Population and discharge capacity



### B) Population density and discharge capacity



### C) Surface area and discharge capacity

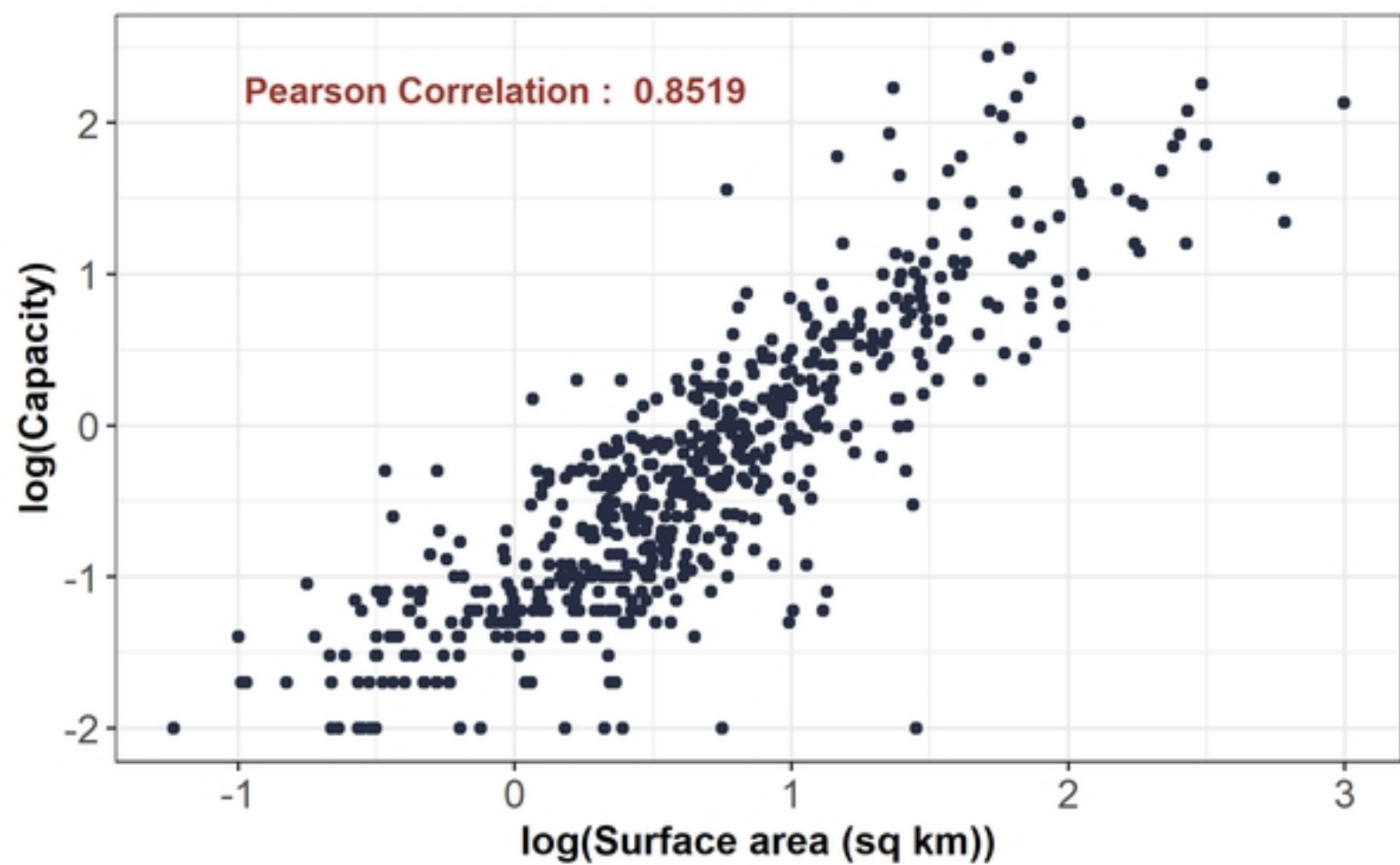


Figure 3