Distribution Agreement

Signature Date

A Mixed-Methods Catchment Area Modelling Approach using Free DEMs to Inform Environmental Surveillance: A Double Case Study Analysis

By

Jamie VanTassell
Master of Public Health
Global Environmental Health

Christine Moe, Ph.D.
Thesis Advisor

A Mixed-Methods Catchment Area Modelling Approach using Free DEMs to Inform Environmental Surveillance: A Double Case Study Analysis

By

Jamie VanTassell
B.S., Georgia Institute of Technology, 2014

Thesis Advisor:
Christine Moe, Ph.D.

An abstract of a thesis submitted to
the Faculty of the Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of Master of Public Health
in Global Environmental Health
2021

Abstract

A Mixed-Methods Catchment Area Modelling Approach using Free DEMs to Inform Environmental Surveillance: A Double Case Study Analysis

By Jamie VanTassell

Environmental surveillance (ES) is an effective epidemiological tool that has successfully been used to monitor poliovirus circulation over the past three decades. ES has also emerged as a low-cost method to detect and estimate incidence of other diseases, such as typhoid fever and COVID-19. A limitation of ES in its current capacity stems from its reliance on sound working knowledge of the sewerage and sanitary system in the setting where it is being used. Without this knowledge, positive detection of a target pathogen in ES samples cannot be traced upstream to the likely source population. Additionally, selecting sites to collect samples for ES is speculative when this knowledge is absent, and samples may not capture the subpopulation of interest and/or may capture overlapping subpopulations. Advances in digital elevation models (DEMs) captured through remote sensing, and their use in hydrologic modeling in ArcGIS, have expanded the potential for ES to be applied to areas where the sewage drainage systems are: 1) not well understood; 2) open to the environment; or 3) gravity-fed. Although a useful tool, historically this approach is less accurate when derived from free, open-access DEM datasets and applied to settings with hydraulic structures. This thesis demonstrates an innovative mixed-methods mapping approach that addresses the current limitations of ES and the use of catchment area delineation for ES using free, open-access data. DEM Reconditioning, a tool in ArcHydro, was applied to improve the accuracy of free DEMs. This methodology was applied to two case studies in unique settings where ES research is underway. The first case study is in Kolkata, India, where ES samples were collected and tested for *Salmonella* (*S.*) Typhi and *S*. Paratyphi A to supplement clinical surveillance of typhoid and paratyphoid fever. A hybrid model was developed that integrates information about hydraulic structures and hydrologic processes that comprise the sewage drainage system in Kolkata. This model was used to estimate the geographic area and population size represented by sewage samples collected from thirteen pumping stations. In the second case study, a catchment area model was developed to identify optimal sampling locations for ES of COVID-19 in Accra, Ghana. This thesis confirmed that a mixed-methods approach can unify free, open-access information and be used to map sewage drainage systems in settings where ES has historically been limited. The Kolkata model demonstrated how this methodology can be used retroactively to inform interpretation of results from ES samples, and it catalyzed discussion about the integrity of the results for settings with flat terrain and extensive artificial canals. The Accra model demonstrated that this process can be used proactively to identify strategic sampling points that capture subpopulations of interest, prevent catchment population overlap, and maximize coverage. When the estimates from the Kolkata and Accra models were compared to watershed-based estimates from a consulting firm for these sites, we concluded that the methodology described in this thesis likely led to more accurate catchment size estimates.

A Mixed-Methods Catchment Area Modelling Approach using Free DEMs to Inform Environmental Surveillance: A Double Case Study Analysis

By

Jamie VanTassell
B.S., Georgia Institute of Technology, 2014

Thesis Committee Chair:
Christine Moe, Ph.D.

A thesis submitted to the Faculty of Rollins School of Public Health of Emory University
in partial fulfilment of the requirements for the degree of Master of Public Health
in Global Environmental Health
2021

Acknowledgements

This thesis would not have been possible without the help and support from mentors, family, and friends.

To my thesis advisor, Dr. Christine Moe, for providing me with invaluable guidance and support every week for many months leading up to the submission of this thesis. Your knowledge and input were critical in shaping this project, and I cannot imagine how it would have turned out without your guidance. Thank you for helping me see the forest for the trees when my mind was in the weeds!

To Dr. Julie Clennon, thank you for patiently answering all my ArcGIS questions during our meetings in the midst of your busy schedule. Your knowledge on all things mapping was an irreplaceable source of help!

To Suraja Raj and Andrew Wang, thank you for reading through sections of my thesis, providing feedback, and facilitating thoughtful discussions that helped shape the trajectory of this project. A special thanks to Suraja for also creating a space for me to vent when I was stressed, always offering an empathetic listening ear, and for the treasured mentorship you have provided throughout my two years at Rollins.

To the members of the environmental data collection team in Kolkata who collected hundreds of samples over the course of several months. This project would not have been possible with the data collected through their tireless efforts.

To my partner, Marshall, and my roommate, Jimmy, for living with me these past two years and being more patient and understanding than I could ever hope for. Thank you for all the times you cooked meals, took out the trash, and mowed the lawn when I was crunched for time. Marshall, thank you for always being there for me when I was emotionally, mentally, and physically drained.

To my dear friends for always sticking by my side despite my mood swings and reclusive tendencies these past few months. I am so grateful for the continued invites to bike rides and hangouts despite all of the times I had to bail or say "no" because I was working on my thesis. Your patience and grace have been more uplifting than I can express in words.

To my family for answering phone calls and listening to me in my times of need. Thank you for always checking in on me, and for not making me feel *too* badly when I was slow at returning phone calls.

And lastly, to my roommate's Golden Retriever, Fitz, who I could always count on to show up during a Zoom call and make loud whining sounds. I do not know how I could have gotten through this thesis without taking breaks to pet you and play tug-of-war while I sat at my computer.

Abbreviations

AMA	Accra Metropolitan Area
Cusecs	Cubic foot per second
DWF	Dry Weather Flow
ES	Environmental Surveillance
DEM	Digital Elevation Model
DSM	Digital Surface Model
GIS	Geographic Information System
KMC	Kolkata Metropolitan Corporation
LMICs	Low-to-Middle Income Countries
PS	Pumping Station
OSM	OpenStreetMap
SHC	Suburban Head Cut
SWF	Town Head Cut
THC	Storm Water Flow
WASH	Water, Sanitation, and Hygiene
WHO	World Health Organization
WWTP	Wastewater Treatment Plant

Glossary of terms

Catchment area	An area that represents the cumulative drainage from sewer and storm water lines, waterways, and surface water runoff to a specific point.
Dry Weather Flow	Flow that occurs during dry conditions. In Kolkata where the sewer and storm water system are combined, dry weather flow (DWF) refers to all flow in the system on a typical day without precipitation.
Delineation	The process of outlining and differentiating adjacent watersheds that do not drain to the same outlet point.
Environmental surveillance	The systematic collection and analysis of samples collected from the environment to measure the presence and/or concentration of a target pathogen in excreta with the objective to monitor infection in a specific population whose excreta is discharged into the environment that is sampled.
Hot spot	An area with elevated incidence or prevalence of disease.
Hydraulic structure	A structure that detains, regulates, or controls the flow of water or wastewater.
Pixel	An individual cell that owns a data value and is stored within a raster of regularly sized cells. Term is used interchangeably with "cell".
Pour point	The point, or outlet, at which water/wastewater from an upstream catchment area drains to. Pour points are determined in ArcGIS using Hydrology Tools by finding the lowest point that a delineated catchment area drains to.
Raster	A matrix, or grid, of regularly sized pixels used to store data.
Sewage drainage system	A system comprised of natural processes and/or man-made infrastructure that influences and directs the flow of sewage throughout a city. The system may be comprised of hydraulic structures that intentionally handle sewage, or it may be driven by natural processes that influence the flow-paths of sewage, such as when sewage discharges into waterways and flows downstream.
Sewerage	The physical facilities or structures (e.g., pipes, pumping stations, treatment/disposal facilities) through which sewage flows.
Storm Water Flow	Flow that includes the effects of precipitation and adds to dry weather flow in a sewage drainage system.

Wastewater The combined flow of blackwater (water that conveys human excreta) and greywater (water used for purposes other than excreta management, such as bathing, washing, etc.) that discharge into the sewage drainage system.

Watershed The area of land consisting of all surface water (precipitation runoff, lakes, waterways, wetlands, etc.) that drains to a shared outlet point. A watershed represents a natural hydrologic entity that captures a specific aerial expanse of land determined by the topography of the terrain. Watersheds can be made up of sub-watersheds and micro-watersheds.

Waterways An umbrella term used to represent all water bodies and features in an area, such as streams, canals, rivers, lakes, ponds, wetlands, etc.

TABLE OF CONTENTS

LITERATURE REVIEW

Infectious Disease Surveillance

Overview

Infectious diseases are a prominent threat to global public health every year. In settings that have not experienced comprehensive, sustainable improvements in water, sanitation, and hygiene (WASH), residents are at an increased risk of morbidity and mortality from infectious diseases due to highly enabled transmission of pathogens in the environment[1]. While low- and middle- income countries (LMICs) bear disproportionately high morbidity and mortality rates for many infectious diseases, the COVID-19 pandemic has demonstrated that the emergence and spread of novel pathogens can have deadly consequences anywhere. With climate change, exponential population growth, rapid urbanization, and increasing antimicrobial resistance all working together to cultivate a more favorable environment for pathogenic organisms, it is critical to systematically monitor the occurrence and spread of infectious diseases through epidemiologic surveillance.

Epidemiologic surveillance is defined by Thacker and Berkelman as the "ongoing systematic collection, analysis, and interpretation of health data essential to the planning, implementation, and evaluation of public practice, closely integrated with the timely dissemination of these data to those who need to know"[2]. Infectious disease surveillance is an effective epidemiological tool that has gained notoriety due to its threefold capabilities: (1) to estimate the current burden of infectious disease within a community; (2) to evaluate spatial and temporal trends of diseases; and (3) early detection of outbreaks and novel pathogens[3]. Infectious disease surveillance is a powerful aid in prevention efforts due to the oftentimes non-uniformity of disease distribution. By identifying where cases are the highest, enhanced monitoring and resources can be funneled into hotspots to provide treatment and curb further spread[4].

There is not a "one size fits all" approach to infectious disease surveillance as a disease monitoring system. A range of different approaches exists with varying advantages and disadvantages[5]. Determining the appropriate approach depends upon multiple factors, such as the setting, available resources, pathogen and disease of interest, and the objectives(s) driving the surveillance[3]. Generally, surveillance approaches can be classified as either active or passive. In passive surveillance systems, a reporting system is established and is maintained by an external agency/institution that is responsible for filing reports on a routine basis. Examples of passive techniques include monitoring of reported diagnostic test results, prescriptions for specific drugs, and hospital admission data[3]. While passive surveillance approaches are relatively inexpensive and straightforward, they are prone to lag and lead to underestimation of the true disease burden due to incompleteness in reporting data and variability in quality amongst providers[5]. Active surveillance systems require public health workers to actively engage with and pursue the data they are interested in retrieving and disseminating[3]. These systems require regular outreach with agencies/institutions to stimulate and maintain the reporting of specific diseases. Advantages of active surveillance systems are that they can validate passive surveillance reports, ensure complete reporting of events, and/or supplement epidemiologic investigations. These systems are often used for short time periods for specific purposes, such as outbreak investigations or monitoring diseases of special interest. Although highly effective and comprehensive, clinical surveillance may miss asymptomatic cases and/or cases that do not seek care or have access to it. In response to these limitations, environmental surveillance (ES), oftentimes referred to as “wastewater-based epidemiology” or “wastewater surveillance”, emerged as a technique to detect silent infections and pathogen circulation.

Environmental Surveillance

Environmental Surveillance (ES) is the heart of this thesis project, and it will be discussed at length to describe how its potential could be expanded for application to multiple infections and a range of settings. ES encompasses the systematic collection of environmental samples that are contaminated with human feces, laboratory analysis for target pathogens, and subsequent interpretation of the results[6,7]. Unlike clinical surveillance which mostly identifies symptomatic cases, ES is able to anonymously and non-invasively expose asymptomatic and undocumented cases within a population through pathogen detection in human sewage samples[6,7]. The undetected cases, or "silent shedders", may be moving throughout the environment and unknowingly transmitting the pathogen to others. ES, therefore, has the potential to provide early warning of pathogen circulation, allowing local health authorities to react accordingly to prevent outbreaks and mitigate further spread.

ES has been used since the 1980s to detect and monitor poliovirus in sewage samples[5,8]. Recognizing its potential to detect silent shedders, World Health Organization (WHO) proposed the use of ES to supplement clinical surveillance of acute flaccid paralysis (AFP) in the Global Polio Eradication Initiative (GPEI)[6,9]. ES has proven to be especially effective in monitoring poliovirus circulation in situations where: 1) vaccine coverage is not comprehensive; 2) previously declared polio-free areas are seeing a resurgence in cases; or 3) contact tracing suggests that future cases are anticipated[10]. By detecting poliovirus in environmental samples, even in areas where cases have yet to be clinically identified, public health workers are able to intervene earlier and funnel resources into high-risk areas.

In addition to its utility in polio surveillance to support the GPEI, ES is useful for monitoring the circulation of other pathogens that are shed in the feces of infected persons and can be detected in wastewater or other environmental samples. Examples of other diseases in which ES has been used to monitor transmission are typhoid fever, cholera, and COVID-19[11–14]. While the overarching concepts of ES are the same, sample collection and laboratory processing and analyses need to be adapted depending on the pathogen of interest. For example, when analyzing environmental samples for *Salmonella enterica* serovar Typhi (*S*. Typhi), which is less persistent in the environment than poliovirus, it is advantageous to sample from locations that reduce the time that sewage has to travel between the source (i.e., location of shedding) and the sampling point[7]. The characteristics of the target pathogen, its persistence in the environment, and detection sensitivity in samples are important factors to consider when designing an ES program and selecting sampling sites.

WHO developed guidelines for ES of poliovirus circulation to provide countries with information on the advantages of ES, methods for selecting sampling sites and laboratory processing, and interpretation and dissemination of the results[10]. The guidelines recommend collecting samples at the inlet of a wastewater treatment plant (WWTP) to capture a large source population (preferably 100,000-300,000 individuals). The guidelines warn that if a formal sewer network does not exist or if the routes of wastewater are not well known, selecting appropriate sampling locations will be difficult and may not be representative of the source population[10]. In theory, this limits the use of ES to areas with formal sewer networks in which toilets are directly connected to the sewer system. In many cities and villages in LMICs, sanitation infrastructure has not yet reached this level of comprehensive connectivity. Oftentimes, sewer networks are informal (i.e., not well documented nor strategically planned) and/or open to the environment (i.e., unprotected drains, canals, or waterways that convey sewage). In such settings, sampling sewage at the inlet of a WWTP (if one exists) may not actually be representative of the source population. This is especially true in areas that use onsite sanitation facilities (e.g., pit latrines or septic tanks) or where open defecation is still practiced because a portion of the population is not contributing feces to the local WWTP.

Understanding the source population that an environmental sample represents, or the “catchment population”, is of the utmost importance in ES. It increases the actionability of a positive lab result by allowing researchers to narrow down the area where infected people are shedding the target pathogen. Additionally, knowing the drainage patterns of wastewater systems and if/how stormwater runoff enters these systems is critical for identifying optimal sampling points and interpreting the results.

Many advances in remote sensing technology and its application to hydrologic modeling have been achieved over the past decade. These advances can be applied to ES to inform interpretation of sample results and strategically identify optimal sites to collect samples from the environment. Learning how to leverage and apply these resources form the building blocks of this thesis work.

Advances in Satellite Imagery and Watershed Modeling

Digital Elevation Models

Digital Elevation Models (DEMs) are the starting point for building the hydrologic models for this thesis project. DEMs are digital representations of the Earth's elevation profile for a specific area[15]. The elevation values are stored in a matrix, and each pixel represents a height above a fixed starting point[16]. DEMs can be generated from land surveys, digital maps, and remote sensing (i.e., satellite imagery)[17]. The accuracy and quality of a DEM is quantified by its resolution, which refers to the area of land that is represented by a single pixel. For example, a 90-meter (90m) resolution DEM is a matrix of 90m-by-90m pixels. Resolution quality can range from 1m (higher quality) to 120m (lower quality). Free, global DEM datasets currently cap at 30m resolution[16]. DEMs can be classified as either digital surface models (DSMs) or digital terrain models (DTMs). DSMs capture the elevation of natural and man-made structures (e.g., vegetation, buildings, etc.) while DTMs represent the "bare-Earth" surface[18]. DTMs are more appropriate to use for modeling hydrologic processes because these models are derived from elevation values and are unobscured by infrastructure and vegetation. Unlike DSM datasets, global DTM datasets are currently not freely available[16–18].

Open access to free global DEM datasets has broken down financial barriers and allowed rapid progress in the application of satellite imagery[17]. The most widely used open-access DEMs were generated from NASA's Shuttle Radar Topography Mission (SRTM) that launched in 2000[16,19]. SRTM data, released globally in 2015, offered the most comprehensive and highest resolution DEMs on a global-scale, with approximately 80% coverage of the Earth's surface and 90m resolution[20]. This had a profound impact on mapping and modeling in LMICs where DEMs were previously unavailable. More recent advances in satellite imagery include the release of 30m SRTM, as well as Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model (ASTER GDEM) released in 2009 by the National Aeronautics and Space Administration (NASA) and the Advanced Land Observing Satellite

Panchromatic Remote-sensing Instrument for Stereo Mapping (ALOS PRISM) released in 2016 by the Japan Aerospace Exploration Agency (JAXA)[21,22].

It is important to select a DEM dataset that most accurately represents the topographic features from which the hydrologic model will be derived as these features directly affect the quality of the model[21]. The appropriate DEM to use in a model will vary depending on the underlying objective, so it is important to research which datasets are most suitable for a particular application. Watershed modeling is relevant to ES as a tool to estimate catchment areas and is the primary focus of this thesis paper. The Gajjar et al. (2018) study concluded that of the free global DEM datasets, the 30m SRTM was the most suitable for watershed modeling, and it was used for this thesis project[23].

Watershed Modeling

The distinction between watersheds and catchment areas varies slightly throughout the literature, and oftentimes these terms are used synonymously. For the purpose of this thesis, the term “watershed” will refer to an area of land for which all surface water (precipitation runoff, lakes, waterways, wetlands, etc.) drains to a shared outlet point. The term “catchment area” will refer to the cumulative drainage from sewer and storm water lines and surface water runoff to a specific point (i.e., a selected sampling location). Therefore, when the term “catchment size” is used, it is referring to the population size (number of people) and geographic area (in square kilometers) contained within a catchment area. The term “sewershed” is intentionally not used in this thesis because it refers to the area of land from which all sewer lines flow to a particular point, and therefore the term is not broad enough to encompass storm water runoff and open channels used for sewage conveyance. The term “sewershed” is appropriate in settings where sewage is contained within a network of closed, mapped sewer lines prior to its disposal.

In watershed modeling, watersheds drain to a common shared outlet, or “pour point”, by gravity (i.e., from higher to lower elevation)[15]. It is helpful to imagine that a droplet of water acts as a marble in a watershed model. If a marble is placed within a watershed, it will either roll towards a depression (i.e., sink) or to a pour point[24]. Through watershed modeling in GIS, watersheds can be delineated to mark the boundaries between watersheds that do not share pour points, and these areas can be used to estimate the catchment areas that drain to a selected sampling point.

The Environmental Systems Research Institute (ESRI) has developed an extensive "Hydrology" toolset within its ArcGIS software that is able to derive hydrologic processes from DEMs. This toolset provides tools that can be executed in a series of iterative steps to generate drainage patterns to delineate watersheds. Arc Hydro, an ArcGIS-based system, contains all of the tools within the Hydrology toolset, and it offers additional functions that are useful for hydrologic modeling. Particularly, the Arc Hydro extension offers an operation called "DEM Reconditioning" that was pivotal for this thesis project[25].

DEM Reconditioning is the process of establishing accurate drainage patterns by "burning" linear features into a raw DEM. These features represent the actual waterways, both artificial and natural, that exist in a study area. The resolution quality of open-access DEMs is oftentimes too low to accurately identify the presence of waterways. Burning these features into an open-access DEM is able to improve the quality of a DEM and the model derived from it. The function uses the AGREE method developed at the University of Texas at Austin in 1997 by reconditioning the surface of an input DEM to agree with input linear features[24,26]. DEM Reconditioning overrides the automatic algorithm built in Arc Hydro that would otherwise calculate the stream network based on the raw DEM. This function essentially recodes the elevation values from the raw DEM along the path of the linear feature, thereby directing the model to recognize the linear feature as the true stream network. To refer back to the marble metaphor, DEM Reconditioning increases the incentive of the marble (i.e., the water droplet) to roll into the burned-in stream network.

The linear features representing the true waterways to be used in DEM Reconditioning can be collected in several ways. The presence of waterways can be physically collected in the field using GPS technology and applications (e.g., mobile data collection using Strava) or extracted remotely from mapping programs like Google Earth or OpenStreetMap (OSM). OSM is an open-access mapping project that is free and user-generated[27]. Waterways extracted from OSM are easy to verify because the features can be compared to waterways in Google Maps.

It is important to note that there are inherent limitations of using free global DEMs in hydrologic modeling. Tavares et al. (2019) summarized many of these and importantly stated that extracting accurate river networks from DEMs was problematic in areas with flat terrain[17]. This is due to the lack of variation in elevation values, which are stored in 30m-by-30m pixels, which makes it difficult for ArcGIS to determine what the direction of flow is because it cannot establish a clear high-to-low drainage route. Identifying a stream network using the Hydrology tools in ArcGIS is preferable for hilly or mountainous environments because the software is able to detect changes in elevation between pixels more easily. However, there are still concerns about the quality of stream networks identified from free DEMs due to the low-resolution quality. The use of DEM Reconditioning in watershed modeling can be utilized as a means to mitigate the issue of imprecise drainage network identification.

Watershed Modeling to Expand the Use of ES

WHO guidelines for ES of poliovirus circulation state that ES is ideal in settings where households have toilets that directly connect to the sewer system because this allows for systematic downstream sample collection[10]. This poses a significant limitation regarding the application of ES to many settings, particularly in LMICs. In areas with underdeveloped sanitation systems and infrastructure, sewage is not contained within a formal network. Instead, human feces may: 1) be contained within pit latrines or other onsite systems that require manual emptying; 2) exist out in the open in areas where open defecation is practiced; or 3) discharge directly into waterways. ES has been shown to be a powerful epidemiologic tool, and there is a need to expand its use to areas without formal sewer networks. These areas typically bear the greatest burden of infectious diseases globally because their informal sanitation systems create enabling conditions for human exposure to fecal contamination in the environment and pathogen transmission.

The application of free global DEM datasets to watershed modeling has great potential for ES of pathogens in various settings. This rationale is behind initiatives by the Bill & Melinda Gates Foundation (BMGF), Environment Systems Research Institute (ESRI), and Novel-T | Innovative Solutions (Novel-T) to generate maps to support active ES programs for global polio eradication[15,28]. The objectives of the maps, which are freely available, are to identify optimal ES sampling locations for polio surveillance and to minimize sampling locations with overlapping catchment areas[29]. The maps include local topography, waterways, and watersheds. However, the waterways and watersheds are not comprehensive for all the settings for which maps have been created, particularly for Kolkata, India. Novel-T openly acknowledges several limitations of the model created for Kolkata: 1) models are derived from 30m DEMs; 2) actual sewer networks are not integrated; and 3) areas with flatter terrain will be less accurate[28].

Other work has been performed to demonstrate the application of watershed modeling to support ES activities. Takane et al. (2016) delineated watersheds using a high resolution (5m) DSM and compared the results to models that used 30m and 90m DEMs[15]. The purpose of the study was to identify optimal sampling sites for ES of polio in cities in Niger and Nigeria and to compare the results generated using different DEM resolutions. The study found that the 5m DSM produced larger watersheds, and therefore the results could help planners identify sampling points that captured a larger population size. The study also concluded that the higher resolution data would be useful in areas with flatter terrain where lower resolution data is unable to accurately capture differences in topography[15]. These are considerable strengths; however, 5m DSM orders cost a minimum of $1,200 (USD) ($3.00 per km^2 and minimum order size of 400 km^2)[30]. This cost poses a financial barrier for ES implementers and researchers who are operating with lean budgets and limited resources. The authors also acknowledge that the application of watershed modeling to ES sample site selection is restricted to areas where the sewer systems are open to the environment and gravity-fed[15]. Sanitation infrastructure may not solely rely on gravity in all settings, and it also may be enclosed and/or located underground. These characteristics cannot be captured by satellite imagery. If the Takane et al. method is applied to a setting where such sanitation infrastructure exists, which is likely in most urban settings, the results will not tell the full story of fecal movement throughout the urban environment.

This thesis project seeks to address the limitations acknowledged by WHO, Novel-T, and the Takane et al. study regarding the scope and applicability of ES. To do so, a methodology that utilizes DEM Reconditioning on free, open-access DEMs and incorporates information about sanitation infrastructure will be performed and analyzed for two settings with different terrain and sewage management systems.

THESIS OBJECTIVES

Building off the previous work of the Takane et al. study and Novel-T maps, DEMs will be used to estimate drainage patterns and delineate catchment areas to inform ES sample site selection and interpretation of ES results. This thesis seeks to address the key limitations acknowledged in previous projects by utilizing free, open-access data and integrating information about sanitation infrastructure into a hybrid (hydrologic-hydraulic) model. Through the proposed methodology, this thesis work will attempt to demonstrate the unmet potential of ES in its current capacity.

There are two ES projects that will be used as case studies to explore this methodology. First, in Kolkata, India, where sampling activities for ES of typhoid and paratyphoid fever were recently completed. Second, in the Accra, Ghana, where sampling activities for ES of COVID-19 is in the planning stages.

This thesis had two overall objectives:

> Objective 1: Address information gaps regarding the movement of human excreta throughout Kolkata and Accra to support the evidence generated by these two unique ES projects.
>
> Objective 2: Develop a mixed-methods approach to build hybrid, context-specific catchment area models using free, open-access data and demonstrate its advantages when applied to ES.

To achieve these objectives, four specific aims were defined:

Aim 1.1: Estimate upstream catchment areas and population sizes for sampled pumping stations in Kolkata, India using freely accessible and open-access data to inform interpretation of ES for *S*. Typhi and *S*. Paratyphi A.

Aim 1.2: Identify optimal wastewater sampling locations in Accra, Ghana using freely accessible and open-access data to support ES of SARS-CoV-2.

Aim 2.1: Develop a step-by-step methodology on how to build an adaptable catchment area model that integrates local sanitation infrastructure and hydrological patterns using freely accessible and open-source data.

Aim 2.1a: Apply the methodology to two contrasting settings to demonstrate its value for ES planning and data interpretation and its potential to be applied to other settings.

The methodology developed in this thesis will demonstrate how it is possible to improve free, open-access DEMs through the process of DEM Reconditioning and how to incorporate relevant, available information on sanitation infrastructure into the model. A detailed, step-by-step description of the methodology will be provided with the intention of making this process transparent and replicable by other researchers (provided in Appendix A). The Kolkata model will be used to improve interpretation of data from previously collected samples for ES for *S.* Typhi and *S*. Paratyphi A by estimating the upstream catchment size (geographic area and population) of sampled pumping stations. The Accra model will be used to identify and propose strategic sampling sites for ES for SARS-CoV-2. Finally, limitations of this methodology will be considered and discussed, and recommendations will be made about how to improve the performance of this methodology in the future.

BACKGROUND

Case Study #1: Kolkata, India

Local Typhoid Fever Burden

S. Typhi and *S*. Paratyphi A, the agents of typhoid and paratyphoid fever, are shed in the feces of infected humans. Deaths from typhoid fever have decreased significantly in the past few decades, from an estimated 600,000 annual deaths in 1984 to between 161,000 and 128,000 deaths in 2017[31,32]. Despite this progress, typhoid fever remains endemic in several regions, with the highest incidence occurring in South-Central and Southeast Asia[33]. The majority of cases occur in children under ten years old. A study conducted in slums in Kolkata estimated an incidence rate of 2 per 1000 population/year for children under five and 5.1 per 1000 populations/year for children under ten years of age[34]. In Kolkata, typhoid fever epidemics have been reported since 1990, and the risk of transmission is heightened by an annual monsoon season that results in flooding in many parts of the city[35,36].

It is difficult to estimate the true burden of typhoid and paratyphoid fever in Kolkata and nationwide due to lack of affordable, accurate diagnostic procedures and inadequate disease-reporting systems[37]. For this reason, the Indian Council of Medical Research-National Institute of Cholera and Enteric Diseases (ICMR-NICED) conducted active community-based surveillance of 6,000 households for clinical symptoms of typhoid from November 2018 to March 2020 in Wards 58 and 59 of Kolkata[7]. To complement this active surveillance, a pilot ES study of typhoid was initiated, and samples were collected from pumping stations and shared household and community toilets weekly[38]. The locations of the pumping station samples will be analyzed using mixed-methods to estimate the upstream catchment areas and population sizes.

State of Sanitation Infrastructure

In 1790, Kolkata (formerly known as "Calcutta") was colonialized and developed into a major river port by the British Empire. This is the reason why, in 1875, Kolkata received a then "state-of-the-art" sewerage network equivalent to the novel system in London. The sewerage network combined storm water, wastewater (which includes blackwater and greywater), and dry weather flow into one system through what is called a “water carriage system”. These systems utilize the buoyant properties of water to convey sewage through pipes and canals[39]. In Kolkata, the main source of drinking water comes from the Hooghly River, which flows north to south and is located west of the city. Tidal gates were installed at the heads of canals to allow water from the Hooghly River to power the water carriage system and divert waste away from the drinking water supply. This design leverages natural water drainage routes, which is generally in a southeasterly direction, from the Hooghly River toward the East Kolkata Wetlands and into the Kulti River, a tidal river that eventually empties into the Bay of Bengal (Figure 1)[36,40].

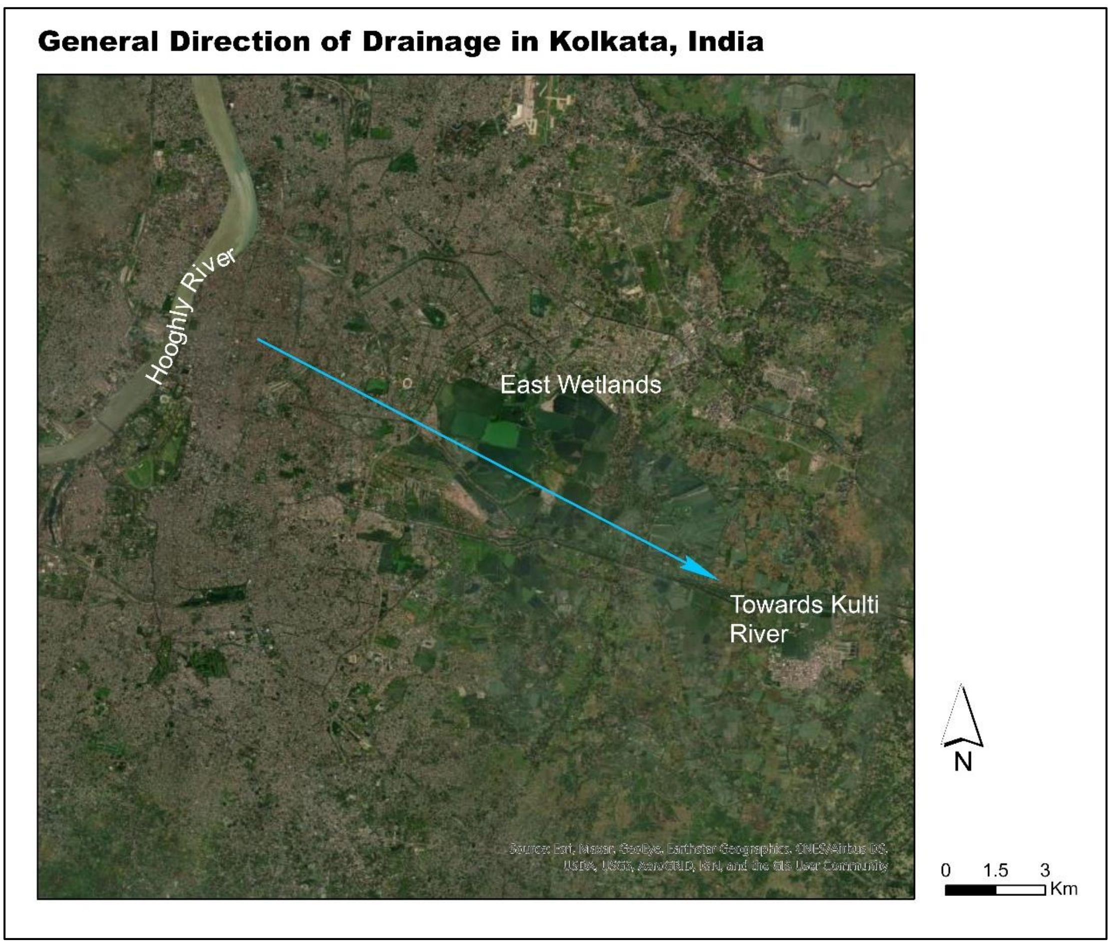


Figure 1. General direction of drainage is in southeasterly direction, from the Hooghly River through the East Kolkata Wetlands towards the Kulti River and, eventually, into the Bay of Bengal.

Kolkata’s sewerage system was initially designed to serve a population of approximately 500,000 people and a total catchment area of 19.1 km^2 for an environment with significantly less paved surfaces[41]. Due to rapid urbanization, the Kolkata Municipal Corporation (KMC) now boasts a population of 4.5 million, with more than one third of its inhabitants living in slums[36,40,42]. This major increase in population adds a significant volume of sewage into the system that was not accounted for during the initial design nearly 150 years ago. In addition, the large increase in paved, impervious surfaces that accompanied urban development to accommodate the higher population decreases water absorption into pervious surfaces such as vegetation and soil[41]. Paved surfaces increase the velocity and volume of rainwater runoff, and this exacerbates flooding in low-lying areas. This is particularly problematic for formal and informal slums which have mostly been constructed in low areas of the city as these were the only remaining areas to be developed[36]. Urban flooding is a recurrent issue in these pockets, especially during and after the monsoon season, and pumping station capacity is insufficient to properly drain the flooded zones. The combination of manufacturing plants that discharge highly toxic pollutants, lack of adequate sanitation infrastructure, and poor drainage present a major environmental hazard that is a threat to public health in Kolkata[36]. The history, complexity, and current issues of the combined sewage and storm water system are the building blocks for developing a catchment area model for Kolkata and its utility for informing ES systems.

Environmental Surveillance Activities

ES for *S*. Typhi and *S*. Paratyphi A was performed by Emory University and NICED in Kolkata between May 2019 to March 2020. Samples were collected from selected pumping stations and shared household and community toilets on a weekly basis either through grab sample or Moore swab method. Vast volumes of combined sewage, storm water flow (SWF), and dry weather flow (DWF) pass through pumping stations, so the samples from these facilities potentially represent large upstream catchment populations. Table 1 provides the GPS location, altitude, and installed pumping capacity for each facility (if it was available[43]) where ES samples were collected. The sampled pumping stations and the boundaries of Wards 58 and 59 are shown in Figure 2. The distribution of selected pumping stations was concentrated in/near these wards to supplement the ongoing active surveillance of clinical symptoms of typhoid at the time. However, ES samples were also collected from pumping stations outside of these wards to collect data from more areas of the city. Rainfall data is also available for the dates in which grab samples were collected and the windows of time in which Moore swabs were in place[44]. Rainfall data is relevant because the pumping stations operate at different capacities after rain events to manage the increased load on the drainage system. Several pumping stations, such as Palmer's Bridge and Ballygunge, actually divert flow to SWF channels when it rains because sewer lines cannot contend with the increased volume in the system[45]. Catchment areas therefore shift depending on the rain conditions, and this affects how samples collected from pumping stations are interpreted.

Table 1. GPS location, altitude, and installed capacity of sampled pumping stations sampled for ES of *S*. Typhi and *S*. Paratyphi A in Kolkata, India

Pumping Station	**Latitude**	**Longitude**	**Altitude (m)**	**Installed Capacity* (cusecs)**
Ambedkar	22.5356091°N	88.3992867°E	14	NA
Baishnabghat	22.4710588°N	88.3924983°E	7	NA
Ballygunge	22.5361962°N	88.3746299°E	3	1,233
Bangur	22.5066961°N	88.3567244°E	12	NA
Chingrighata	22.5574774°N	88.3920952°E	7	100
Cossipore	22.6060572°N	88.3734558°E	7	NA
Dhapa Lock	22.5576242°N	88.4110281°E	7	480
Duttabagan	22.6078184°N	88.3923887°E	5	48
Jorabridge	22.4956886°N	88.3913613°E	7	NA
Kulia Tangra	22.5580645°N	88.3894534°E	8	40
Pagladanga	22.5561565°N	88.3959111°E	7	48
Palmer's Bridge	22.5632013°N	88.3777120°E	7	1,184
Topsia	22.5408928°N	88.3868116°E	7	65

NA = Not Available
*Installed capacities from Government of West Bengal, Irrigation & Waterways Department[43]

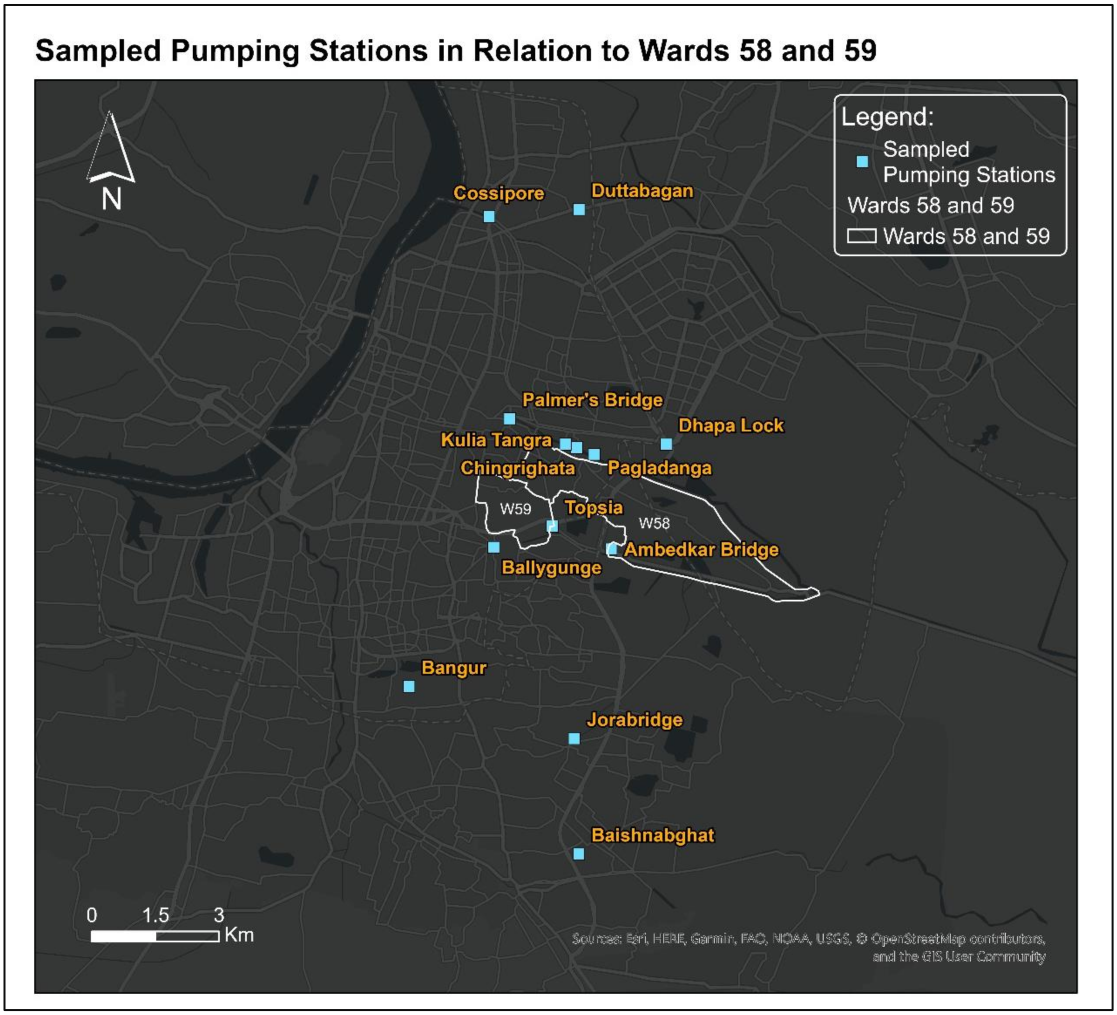


Figure 2. Locations of pumping stations sampled for ES for *S*. Typhi and *S*. Paratyphi A in relation to Wards 58 and 59, where active surveillance of clinical symptoms of typhoid fever was conducted. Selected pumping stations for ES were concentrated in/near these wards to supplement active surveillance.

Case Study #2: Greater Accra Metropolitan Area, Ghana

Local COVID-19 Burden

COVID-19 has disproportionately affected the Greater Accra region compared to the rest of Ghana. More than 56 percent of the nation's diagnosed cases have occurred in the Greater Accra region, with the second highest region, Ashanti, experiencing more than 34,000 *less* cases[1] despite having a larger population[46]. This disparity is significant and highlights the necessity of increasing resources and monitoring to control further spread of COVID-19. Organized efforts have been made to control the spread of COVID-19 in Greater Accra, such as mandating the use of face coverings in public places and implementing a system of contact tracing[47]. Despite these increased efforts, compliance with COVID-19 prevention measures remains low, and the Greater Accra region continues to bear the highest number of recognized active cases[46,48].

State of Sanitation Infrastructure

The Greater Accra District has experienced rapid urbanization over the past three decades, which is associated with better national economic performance and job creation. However, the sanitation infrastructure has not kept pace with the rapid population growth and economic prosperity. Due to the sprawling nature of Greater Accra, which is continuously widening with urbanization, it is difficult to expand basic services and infrastructure to reach newly developed areas. Compared to other Ghanian cities, Accra saw the steepest decline in access to piped water from 2000 to 2010[49]. The local government is limited in its capacity to increase provision while maintaining and managing its aging, inadequate sanitation infrastructure[50]. A World Bank report identified poor sanitation as one of the city's key stressors and highlighted its contribution to exacerbated flooding outcomes and annual cholera outbreaks[51]. It is interesting to note that although Accra was formerly a British colony like Kolkata, it did not receive an elaborate sewerage system like Kolkata did.

Sewage is often discharged into open drains that were constructed to drain storm water in Accra[52]. As shown in Figure 3, the open drains discharge into larger waterways that flow in a southerly direction to the ocean. Maps of open drains were not available for all of Accra in OSM, but the snapshot of open drains shown in the northeast corner in Figure 3 shows how interwoven these structures are in the urban

[1]Statistic to-date as of April 1st, 2021 from the Ghana Health Service website[46]

landscape and how they discharge into waterways that lead to the ocean. Open drains are prone to overflow during the spring and fall rainy seasons and release fecal contamination into the surrounding environment, putting residents at risk of exposure to fecal contamination[53,54]. Open defecation has been declining in recent years, but it has not yet been eliminated and still poses a risk to human health.

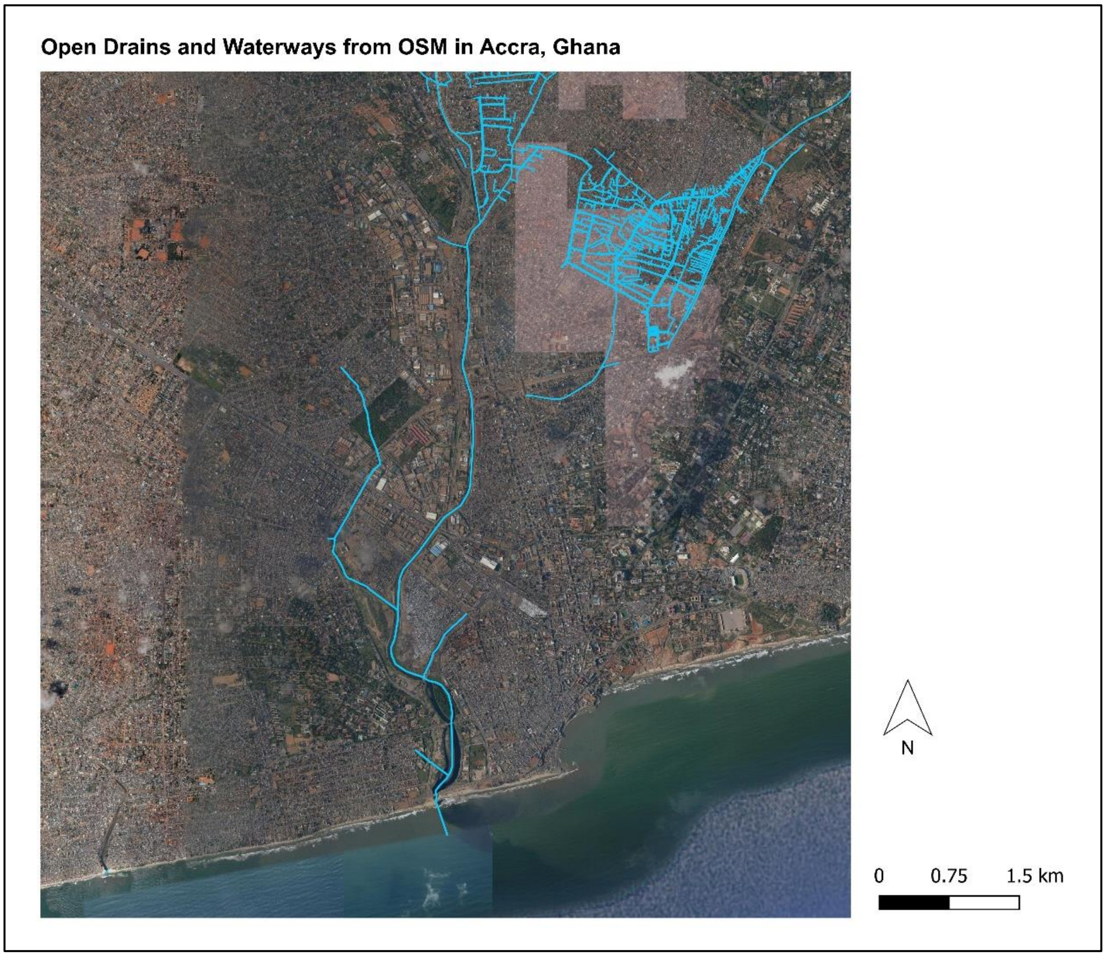


Figure 3. Open Drains and Waterways from OSM for Accra, Ghana. The open drains shown in the northeast corner of the map were not available for all of Accra from OpenStreetMap (OSM). However, providing a snapshot of these drains shows how interwoven they are in the urban landscape and how they drain into larger waterways that ultimately lead to the ocean.

Environmental Surveillance Activities

As of April 2021, ES samples have been collected in the Tema and Korle Klottey districts within the Greater Accra Region, but samples have not yet been collected in the AMA. The outputs of the catchment area model from this thesis project will help identify strategic sampling points in the AMA prior to sample collection for ES of COVID-19[7]. Importantly, the model is able to identify points downstream of open drains that convey excreta from neighborhoods since these flow by gravity into larger waterways. In addition, this proactive approach to identify strategic sampling points enables ES implementers to prioritize certain areas in the AMA that are more vulnerable to COVID-19 infection and ensure that these areas are captured by sampling locations. Examples of criteria for identifying vulnerable neighborhoods include higher population density, lower socioeconomic status (SES), limited sanitation infrastructure/services coverage, higher rates of open defecation (indicating greater poverty and use of WASH facilities), and experiencing a history of cholera outbreaks. Once sampling activities for ES commence in the AMA, the catchment area model may help programmers trace samples that are positive for SARS-CoV-2 RNA upstream to locate cases. This is useful to provide early warning of surges in COVID-19 infection and help prioritize resources into hot zones. The model is also able to identify areas of the city where pathogens would need to travel a considerable distance before arriving at the sample location, which may result in false negative ES results and underestimates or under-recognition of the COVID-19 burden from these areas.

MANUSCRIPT

Abstract

Environmental surveillance (ES) is an effective epidemiological tool that has successfully been used to monitor poliovirus circulation over the past three decades. ES has also emerged as a low-cost method to detect and estimate incidence of other diseases, such as typhoid fever and COVID-19. A limitation of ES in its current capacity stems from its reliance on sound working knowledge of the sewerage and sanitary system in the setting where it is being used. Without this knowledge, positive detection of a target pathogen in ES samples cannot be traced upstream to the likely source population. Additionally, selecting sites to collect samples for ES is speculative when this knowledge is absent, and samples may not capture the subpopulation of interest and/or may capture overlapping subpopulations. Advances in digital elevation models (DEMs) captured through remote sensing, and their use in hydrologic modeling in ArcGIS, have expanded the potential for ES to be applied to areas where the sewage drainage systems are: 1) not well understood; 2) open to the environment; or 3) gravity-fed. Although a useful tool, historically this approach is less accurate when derived from free, open-access DEM datasets and applied to settings with hydraulic structures. This thesis demonstrates an innovative mixed-methods mapping approach that addresses the current limitations of ES and the use of catchment area delineation for ES using free, open-access data. DEM Reconditioning, a tool in ArcHydro, was applied to improve the accuracy of free DEMs. This methodology was applied to two case studies in unique settings where ES research is underway. The first case study is in Kolkata, India, where ES samples were collected and tested for *Salmonella* (*S.*) Typhi and *S.* Paratyphi A to supplement clinical surveillance of typhoid and paratyphoid fever. A hybrid model was developed that integrates information about hydraulic structures and hydrologic processes that comprise the sewage drainage system in Kolkata. This model was used to estimate the geographic area and population size represented by sewage samples collected from thirteen pumping stations. In the second case study, a catchment area model was developed to identify optimal sampling locations for ES of COVID-19 in Accra, Ghana. This thesis confirmed that a mixed-methods

approach can unify free, open-access information and be used to map sewage drainage systems in settings where ES has historically been limited. The Kolkata model demonstrated how this methodology can be used retroactively to inform interpretation of results from ES samples, and it catalyzed discussion about the integrity of the results for settings with flat terrain and extensive artificial canals. The Accra model demonstrated that this process can be used proactively to identify strategic sampling points that capture subpopulations of interest, prevent catchment population overlap, and maximize coverage. When the estimates from the Kolkata and Accra models were compared to watershed-based estimates from a consulting firm for these sites, we concluded that the methodology described in this thesis likely led to more accurate catchment size estimates.

Introduction

Environmental surveillance (ES) for enteric infections in which the pathogens are shed in feces is a growing trend. Successful applications of ES for poliovirus1 demonstrated that ES can act as a sensitive system for early detection and monitoring of polio. It is now being adapted and used for surveillance of other pathogens in wastewater such as *S*. Typhi and *S*. Paratyphi A and SARS-CoV-2[7,14,55]. These diseases are often driven by asymptomatic or mild symptomatic carriers. Therefore, clinical surveillance can underestimate the true burden of disease in a community and be ineffective as an early warning system[6,7].

One constraint of the ES method is that it relies on sound, working knowledge of local sewer networks and drainage systems to be actionable in a given setting. Without this knowledge, it is not possible to link pathogens detected in wastewater to the source population. In many cities in LMICs, sewer networks may be informal, unmapped (or not possible to access maps), and/or open to the environment (e.g., discharged to open drains and canals). In such cases, the use of ES is complicated because the source (i.e., the infected person(s) shedding the pathogen) cannot be readily identified and connected with pathogen detection in wastewater samples. The term used to denote the population that contributes to a sample point or is served by a facility is oftentimes referred to as a catchment population[7,15]. Throughout this paper, the term "catchment area" will be used to represent the cumulative drainage from sewer and storm water lines and surface water runoff to a selected sampling location.

The Takane et al. (2016) study demonstrated the use of satellite imagery to identify optimal locations to collect samples for ES of polio[15]. To do this, a high-resolution DSM was used to derive hydrologic characteristics and delineate watersheds. Watersheds drain to shared outlets, or "pour points", and these points may serve as strategic sampling points because they capture wastewater from a large geographic areas and population sizes while minimizing overlap between catchment areas. This study importantly laid the foundation for applying watershed delineation to sample site selection for ES, but it had several limitations. The study was able to purchase and utilize a 5m mesh DSM, and the cost of this data is a minimum of $1,200 ($3.00 per km, minimum order of 400 sq. km.)[15,30]. Additionally, the study was only applied to areas where wastewater flow was conveyed through open drains and canals and driven by gravity (i.e., no underground pipes or pumping stations) and conceded this as a limitation of the method.

While the concepts used in the Takane et al. study are similar to those in this study, the cost requirements and limited applicability acknowledged by the paper inspired further research into these challenges. There are two main research questions motivating this project:

> Research Question 1: During ES for pathogens in wastewater, how do researchers interpret what the samples represent (i.e., the source population) when sewage conveyance systems and wastewater flow patterns are unknown, complex, and/or open to the environment?

> Research Question 2: How can free, open-access DEMs be applied (in place of costly satellite imagery) to identify strategic sampling locations for ES?

The Center for Global Safe WASH (CGSW) at Emory University is involved in two ES projects at this time. First, ES for *S*. Typhi and *S*. Paratyphi A in selected wards in Kolkata, India was performed in partnership with the Indian Council of Medical Research-National Institute of Cholera and Enteric

Diseases (ICMR-NICED). This was conducted between May 2019 through March 2021, and the results are currently being analyzed. Typhoid fever is endemic in Kolkata, and epidemics have been reported since 1990, oftentimes following the annual monsoon season that creates city-wide flooding[35,36]. This research is important for estimating the local burden of typhoid fever and will be used to supplement clinical surveillance data. However, the sewer connectivity and upstream catchment areas of the environmental sampling points are not well understood, and therefore interpreting how the ES results reflect spatial and temporal patterns of typhoid prevalence is challenging.

Second, the CGSW is currently implementing ES of COVID-19 in neighborhoods in the Greater Accra District, Ghana by testing wastewater samples for presence of SARS-CoV-2. Sample collection has commenced in two neighborhoods, but samples have not yet been collected in the Accra Metropolitan Area (AMA). The results of diagnostic testing indicate that the Greater Accra region has been hit the hardest by COVID-19 compared to other regions in Ghana, bearing more than 55 percent of the nation's cases despite having a smaller population than the Ashanti region[46]. This disproportionate burden highlights the need for increased attention and resources for the Greater Accra District. Sampling activities in the AMA have not been initiated at present, and this offers the opportunity to explore the use of free DEM data to identify strategic sampling locations for ES.

This thesis had two overall objectives:

> Objective 1: Address information gaps regarding the movement of human excreta throughout Kolkata and Accra to support the evidence generated by these two unique ES projects.
>
> Objective 2: Develop a mixed-methods approach to build hybrid, context-specific catchment area models using free, open-access data and demonstrate its advantages when applied to ES.

To achieve these objectives, four specific aims were defined:

Aim 1.1: Estimate upstream catchment areas and population sizes for sampled pumping stations in Kolkata, India using freely accessible and open-access data to inform interpretation of ES for *S*. Typhi and *S*. Paratyphi A.

Aim 1.2: Identify optimal wastewater sampling locations in Accra, Ghana using freely accessible and open-access data to support ES of SARS-CoV-2.

Aim 2.1: Develop a step-by-step methodology on how to build an adaptable catchment area model that integrates local sanitation infrastructure and hydrological patterns using freely accessible and open-source data.

Aim 2.1a: Apply the methodology to two contrasting settings to demonstrate its value for ES planning and data interpretation and its potential to be applied to other settings.

Methods

Case Study #1: Kolkata, India

ES Study

ES for *S*. Typhi and *S*. Paratyphi A was performed in Kolkata between May 2019 to March 2020. Samples were collected from pumping stations and shared toilets weekly either through grab sample or Moore swab method. Thirteen pumping stations were sampled during the project.

Investigative Research

An exhaustive online search was performed to gain knowledge about Kolkata's complex sanitation landscape. The sources of information included textbooks, historical city planning reports, records of upgrades and projects, scientific literature, government websites, surveyed performed at pumping stations, and news and blog articles. An exhaustive search was necessary to conduct for Kolkata due to the drainage system's vastness, complexity, and history of patchwork upgrades. The thirteen sampled pumping stations were extensively researched to better understand how these facilities dictated and altered the flow of sewage throughout the city. A consistent finding across the various sources was that the general direction of water flow was eastward across Kolkata (Figure 4). Other important findings from this search used to inform the model included: maps and descriptions of drainage basins[45], diagrams of underground sewer and storm water lines for three drainage systems[44], and installed pumping station capacity numbers[43].

Building the Model

The Coordinate Reference (CRS) used for this project was “WGS 1984 / UTM Zone 45N” due to Kolkata’s location between 84°E and 90°E in the northern hemisphere above 84°N[56]. Sewerage diagrams depicting underground connections to three of the thirteen pumping stations were found and were manually digitized in ArcGIS. Waterways were then extracted by running a query in OpenStreetMap’s data-mining tool “OverPass Turbo”. Several waterway segments were missing or incomplete, and these features were manually edited by tracing streams in Google Maps that were not captured by the query. The result was a complete network of open canals, channels, and streams within the Kolkata Metropolitan Corporation (KMC). The underground sewer and storm water pipes, open waterways, and sampled pumping stations were integrated to piece together the puzzle of how the sewage drainage system is connected in Kolkata (Figure 4).

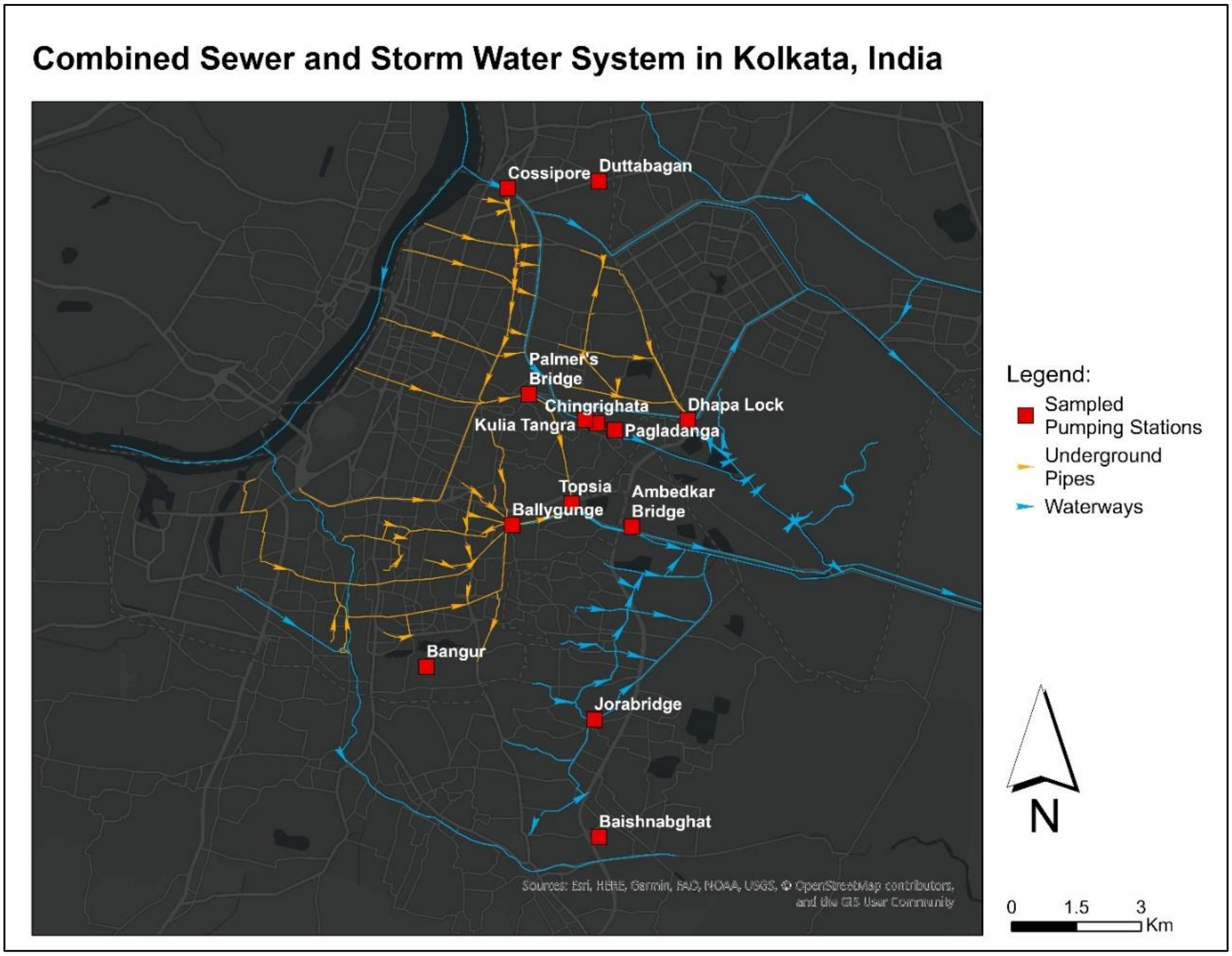


Figure 4. Connections between underground sewer and storm water pipes, open waterways, and sampled pumping stations were pieced together to better understand the movement of sewage and storm water throughout Kolkata, India.

The DEM for Kolkata was downloaded from USGS EarthExplorer and added to the map. The DEM Reconditioning tool was then used to burn the three waterway rasters into the DEM. This process was performed separately and iteratively for each raster (i.e., the output from the previous operation was used as the input in the next operation). The values shown in Table 2 were used as the inputs for the DEM Reconditioning tool. The purpose of this process was to lessen the impact of small and medium waterways on the DEM to create a reconditioned raster with accurate drainage patterns. The values were significantly smaller compared to the default settings for the field “Sharp drop in Z units” because

Kolkata’s terrain is very flat and close to sea-level. Due to these characteristics, using a large value for the sharp drop field would over-emphasize drainage into these canals and warp true drainage patterns. The effects of DEM Reconditioning are depicted in Figure 5.

Table 2. Values input in fields in the DEM Reconditioning tool in Arc Hydro Pro Tools compared to the default settings.

	Number of cells for stream buffer	Smooth drop in Z units	Sharp drop in Z units
Small waterways	1	1	10
Medium waterways	2	5	10
Large waterways	5	3	10
Default Settings	5	10	1,000

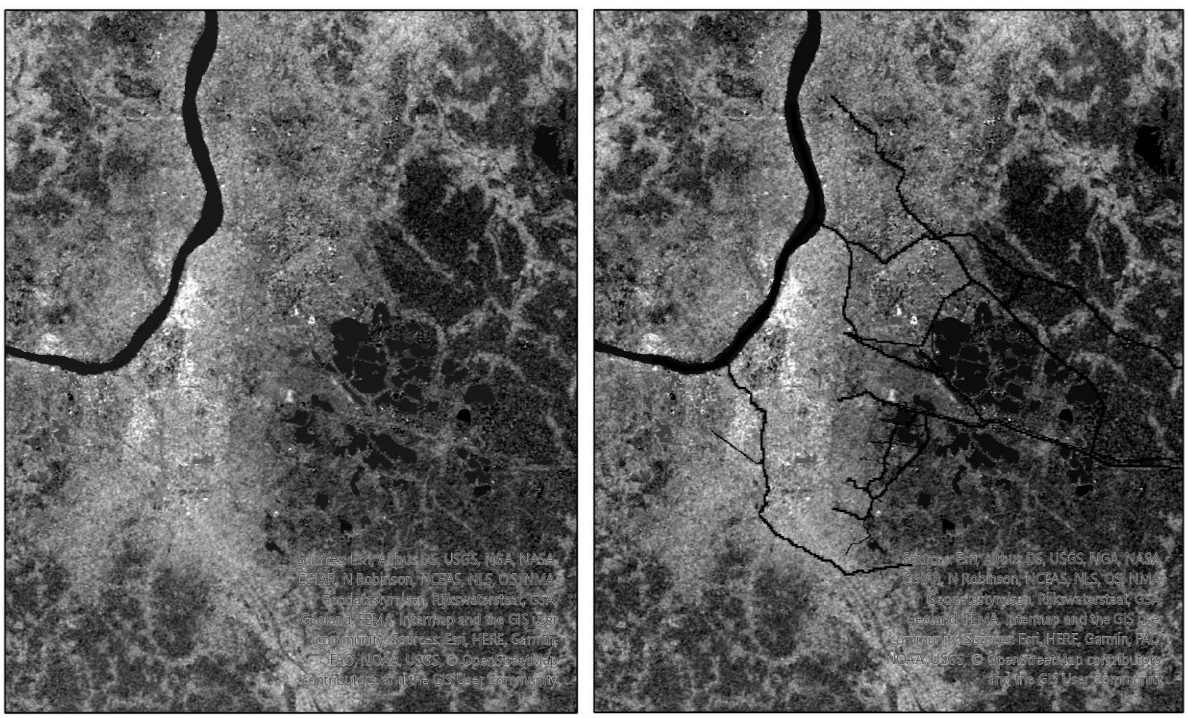


Figure 5. Before DEM Reconditioning tool was applied (left) and after DEM Reconditioning tool was applied. The black lines burned into the DEM in the reconditioned map represent artificial waterways (i.e., man-made canals) that were not captured through satellite imagery.

The next sequence of steps was performed using the Arc Hydro extension in ArcGIS Pro[25,26]. The reconditioned DEM was run through the "Fill Sinks" tool to massage out erroneous sinks and depressions. The output was then processed by the "Flow Direction" tool to chart the direction of water flow leaving each cell using an eight-direction flow model. In this model, each cell is assigned a direction that represents one of eight possible directions that water would take as it exits a cell (1=East, 2=Southeast, 4=South, 8=Southwest, 16=West, 32=Northwest, and 64=North). These values are assigned based on each cell's elevation value relative to its neighbor, and the direction of flow will point towards the cell's neighbor with the lowest elevation value. The output from this operation was then input into the "Flow Accumulation" tool. This function tallies how many cells point towards, or accumulate, into each cell. Next, the output was run through the "Raster Calculator" tool to specify how many accumulated cells represent a likely stream network. This tool assigns a value of "1" to cells above a threshold value and "0" to cells below this value. The threshold was set at 1,000 cells because the output most closely resembled the waterways in OSM and Google Maps. The output from this operation was then run through the "Stream Segmentation" tool which creates a unique feature for each segment of the stream network. A catchment area was assigned to each stream segment by processing this output in the "Catchment Grid Delineation" tool. Lastly, a polygon layer was created by inserting the catchment grid into the tool "Catchment Polygon Processing" tool.

Integrating Population and Demographic Data

A shapefile with population data from the India Census of 2011 organized by wards was found within the ArcGIS online gallery[57]. Figure 6 shows categorization of wards by population size using the Natural Breaks (Jenks) classification.

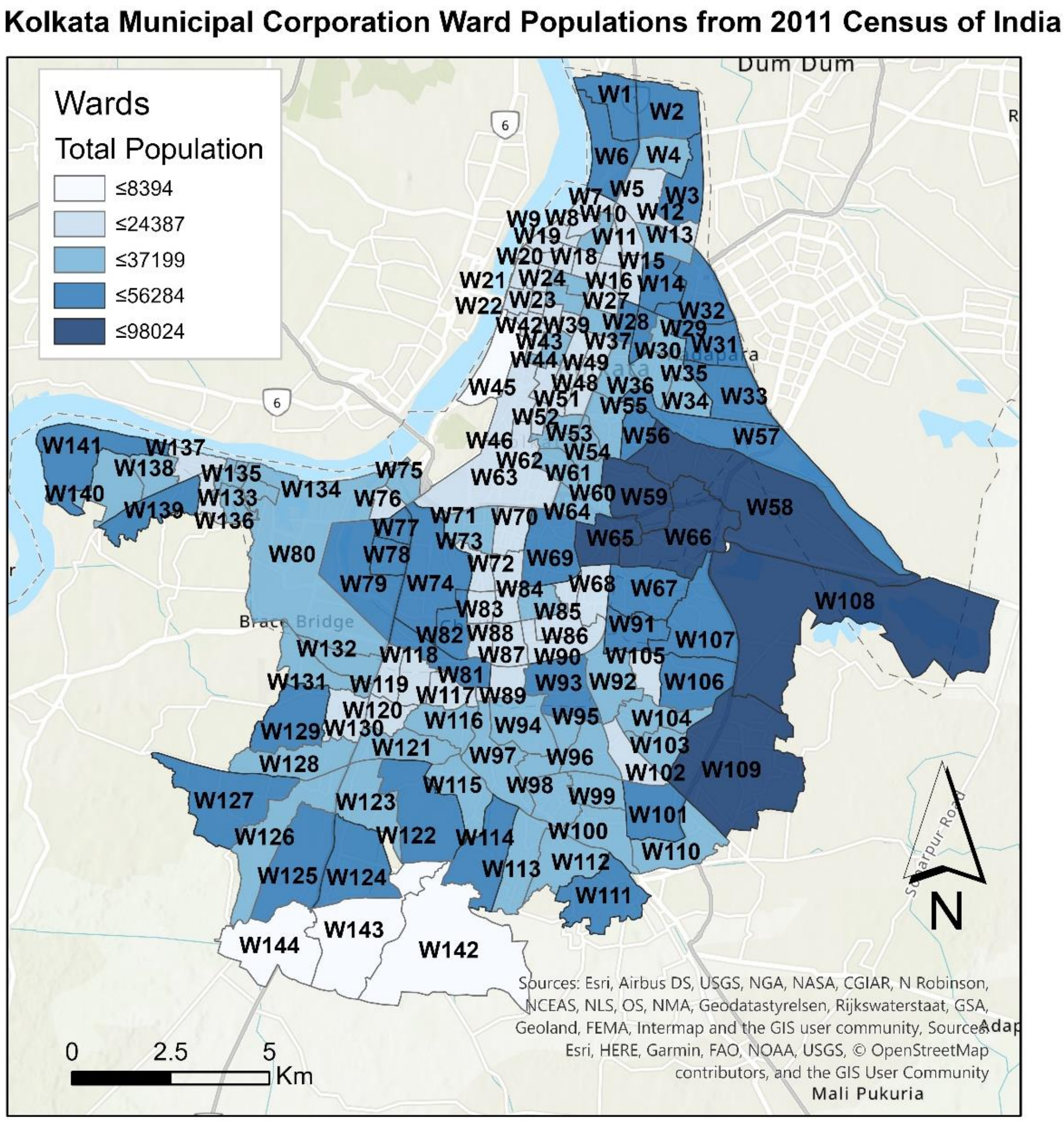


Figure 6. Total population per ward from 2011 Census of India[42] using Natural Breaks (Jenks) classification

Estimating Catchment Areas of Sampled Pumping Stations

Due to the plethora of studies and information on the history of Kolkata's sewerage and storm water system, many factors went into the decision-making process for estimating the population sizes of the catchment areas. A process flow diagram was created to illustrate this process (Figure 7). The methodology for estimating the catchment areas of sampled pumping stations and their population sizes operates under two key assumptions: 1) population density is consistent throughout a ward; 2) unknown underground sewer and storm water lines follow the drainage patterns identified in Arc Hydro in areas where no information was available about these connections.

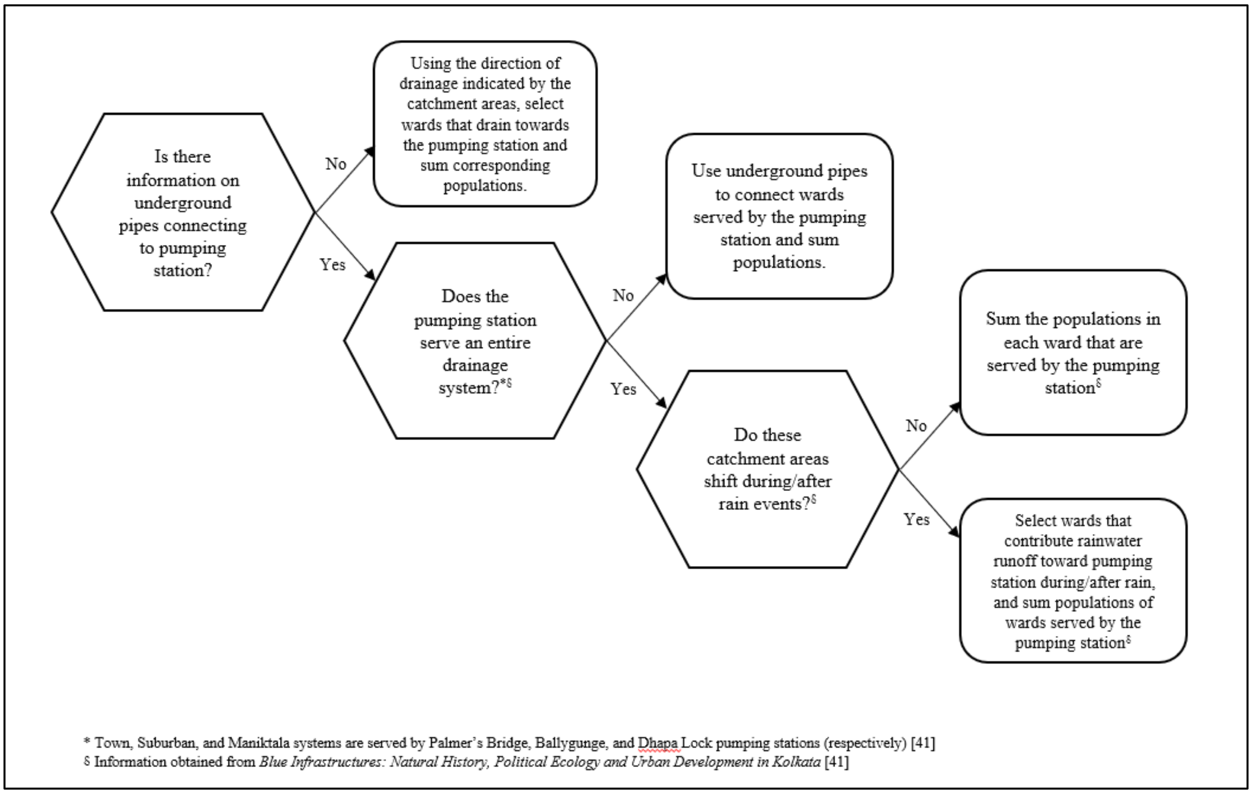


Figure 7. Process Flow Diagram for Estimating Catchment for Sampled Pumping Stations in Kolkata, India

The Duttabagan pumping station (PS) will be discussed in detail to provide an example of how the majority of catchment area population sizes were estimated. The justification and rationale behind population estimates for all of the sampled pumping stations is provided in Appendix B.

The was no information available about the underground pipe network that connects sewer and storm water to the Duttabagan PS, and it is therefore in the first “No” box in the process flow diagram (Figure 7). The delineated catchment grid shown in Figure 8a suggests that the catchment area that houses the

Duttabagan PS drains southward. A polygon was drawn to represent the smaller catchment area that drains toward the facility (Figure 8b). This process revealed that partial areas of Wards 2, 3, and 4 drain to the Duttabagan PS. The "Tabulate Intersection" tool was then applied to determine the percentage of area that Wards 2, 3, and 4 overlap with the drawn polygon. The percentages were multiplied by the total population in each ward and summed to calculate the total population size of the catchment area.

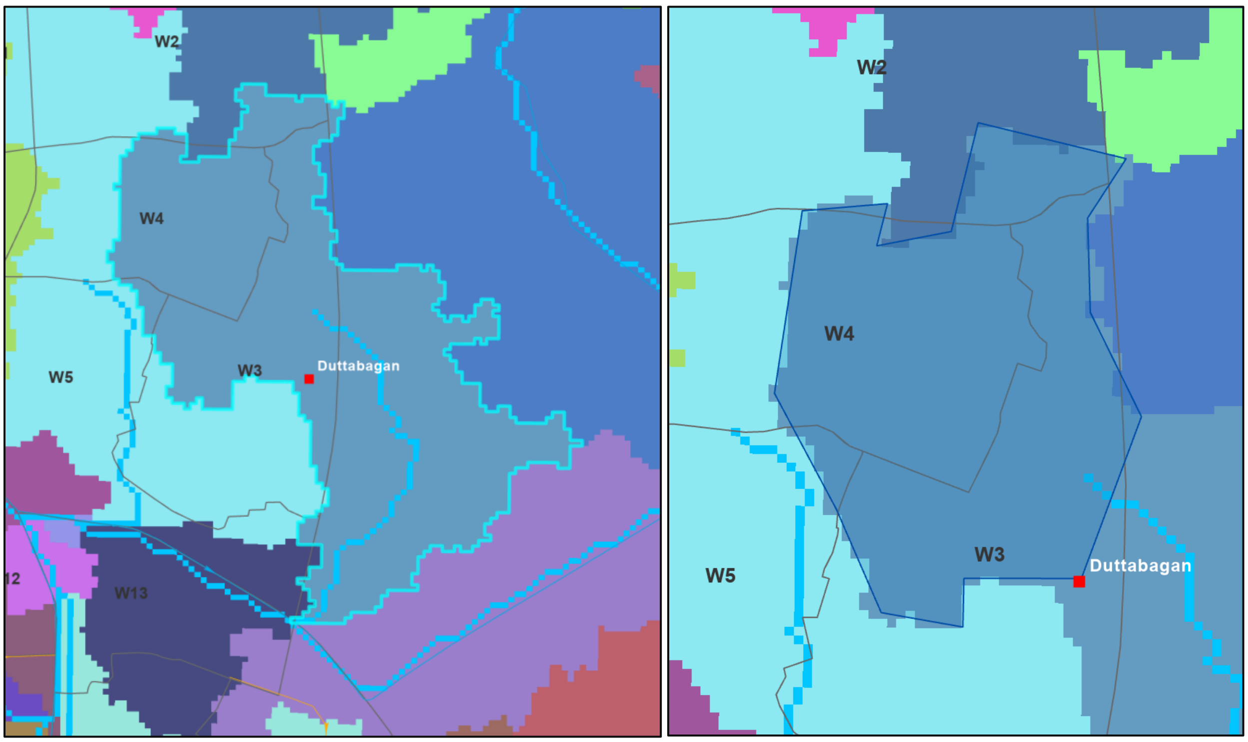


Figure 8a (left). Watershed in which Duttabagan PS is situated. **Figure 8b** (right). Perceived zone of catchment area that drains to the Duttabagan PS.

For those who are interested in replicating this study or applying this methodology, a step-by-step tutorial is provided in Appendix A. The catchment area calculations for each sampled pumping station in Kolkata are provided in Appendix B.

Case Study #2: Accra, Ghana

ES Background

As of April 2021, ES samples have been collected in the Tema and Korle Klottey districts within the Greater Accra Region, but samples have not been collected in AMA yet. Results from the Accra model will be used to recommend strategic sites to collect ES samples from waterways.

Investigative Research

The sewer system and terrain of Accra is very different from Kolkata, and this affected how the Accra model was built. Research was first performed to gain an understanding of the sanitation infrastructure in AMA, but it was a less robust process compared to the research performed for Kolkata because the sewer system in Accra has less hydraulic structures that disrupt the flow of wastewater. After conversations with local partners in the TREND Group (a local environmental consulting group that specializes in sanitation), no pumping stations were added to the model because the few existing facilities were nonfunctional. In addition, underground sewer and storm water lines were not incorporated because these structures are not common in the AMA. Owing to the city's primary reliance on vehicle transport of excreta for treatment/disposal and the common practice of discharging excreta into open drains, it was assumed that these structures would have a negligible impact in the model.

There is a vast network of open drains throughout Accra that are intended to collect and drain storm water, although wastewater and solid waste are often deposited into these drains[52,58]. Accra is a hilly, coastal city, and as was previously shown in Figure 3, open drains connect to larger waterways which drain by gravity to the ocean. Although maps of open drains were not available for all of the neighborhoods in Accra, a hydrologic model is able to capture these complex networks because open drains flow by gravity into larger waterways. Therefore, strategic sampling points along the larger waterways capture areas from which excreta is discharged into open drains.

Open defecation has been declining in recent years, but it has not been completely eliminated. This is relevant to the Accra model because uncontained feces in the environment may follow DEM-derived drainage routes after heavy rain events (assuming there are limited barriers), and once excreta are washed into waterways, it can be captured by ES samples. Pit latrines and septic tanks are commonly used to hold sewage in Accra, and the contents are emptied and trucked offsite to facilities for disposal and/or treatment. It is important to note that the watershed model developed for Accra was not able to capture the movement of sewage that was transported from holding reservoirs to offsite facilities.

Building the Model

The CRS used for the Accra model was "WGS 1984 / UTM Zone 30N" due to its location between 6°W and 0°W in the northern hemisphere above 84°N[59]. Similar to Kolkata, the DEM for Accra was downloaded from USGS EarthExplorer and waterways shapefiles were extracted in OSM using Overpass Turbo. The waterways shapefile was edited to match streams in Google Maps and was converted into a raster. The raster was input in the DEM Reconditioning tool (due to there being less variation in stream widths, there was no need to classify waterways into categories as was done for Kolkata). The default settings in the DEM Reconditioning tool were used because these settings were considered appropriate for Accra's hilly terrain. The next series of steps, from "Fill Sinks" to "Catchment Polygon Processing", mirrored the methods used for Kolkata.

Integrating Population and Demographic Data

A shapefile of 2010 census data for neighborhoods in Accra was provided by local partners in the TREND Group in Ghana[60]. The data stored in the shapefile included population counts and summary statistics of sanitation infrastructure and behaviors (e.g., open defecation rate) at the time the survey data was collected. Figure 9 shows population counts of neighborhoods categorized using the Natural Breaks (Jenks) classification.

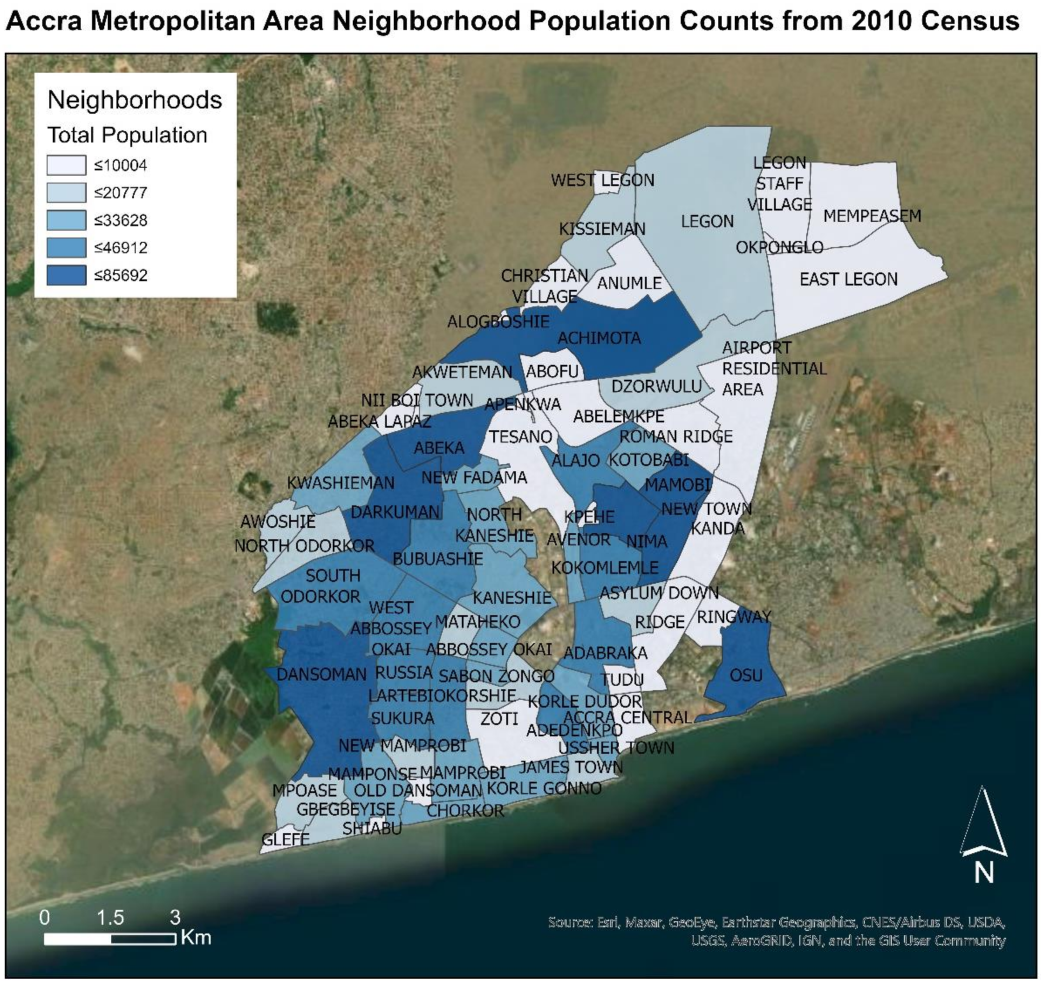


Figure 9. Total population per neighborhood from 2010 Census using Natural Breaks (Jenks) classification (shapefile used to create map provided by TREND Group in Ghana).

Estimating Catchment Area and Population Size

This procedure mimics the methods used for Kolkata in which the intersection between delineated watersheds and neighborhood areas was calculated to estimate the catchment population size of strategic sampling points. Similar to Kolkata, it also assumes that population density is consistent at the neighborhood-level. Since no samples have been collected yet in Accra, pour points that captured a wide geographic catchment area were selected as optimal positions to collect ES samples. Priority was given to pour points that captured areas that met certain criteria, such as if they contained medium-to-high population size, higher rates of open defecation, low SES, and/or previous occurrence of cholera outbreaks[61]. These factors were considered relevant to WASH infrastructure, behaviors, and COVID-19 transmission. Comprehensive geographic coverage was also a factor in selecting locations because samples will ideally capture a majority of the population living in AMA. Minimizing overlap between catchment areas was also deemed a priority in sample site selection, and pour points were selected that each represented a unique geographic area. Pour points were excluded if their upstream catchment areas were primarily located outside the boundary of the AMA. Once the optimal sampling locations were selected, the "Tabulate Intersection" tool was used to determine the percentage of area that each neighborhood shares with the upstream catchment area of the pour point.

Results

Case Study #1: Kolkata, India

The catchment area size estimates (geographic area and population) for all the pumping stations sampled and tested for *S*. Typhi and *S*. Paratyphi A are shown in Table 3. Geographic size of catchment areas ranged from 0.06 – 20.42 km^2, and the population estimates of catchment areas ranged from approximately 1,493 – 890,681 people. The average catchment area was 3.69 km^2 and included approximately 134,050 people. The standard deviations for the geographic area and population estimates of the catchment areas was 6.04 km^2 and 252,990 people, respectively. These summary statistics are reported in Table 4. The Pearson correlation test indicated a strong positive linear relationship between geographic area and estimated population size ($r = 0.99$), which is to be expected as larger areas will oftentimes contain more people in urban settings.

In Table 3, the catchment population estimates were categorized by their upper order of magnitude and color-coded to illustrate the distribution of samples collected from pumping stations with population estimates that are: 1) <10,000; 2) 10,001 – 100,000; and 3) 100,001 – 1,000,000 people. For the first category (color code = yellow), 57 total wastewater samples were collected from five different pumping stations. For the second category (color code = blue), 101 total samples were collected from five different pumping stations. Finally, for the last category (color code = green), 74 total samples were collected from three different pumping stations. These results indicate that the samples captured a wide distribution of catchment population sizes, and this finding has important implications for the interpretation of the ES results that will be examined in the Discussion section.

Table 3. Number of samples collected and estimated geographic catchment and population size for sampled pumping stations in Kolkata.

Pumping Station	No. Samples Collected	Geographic Catchment Area (sq km)	Population Size Estimate
Kulia Tangra	20	0.12	1,493
Bangur	5	0.06	1,525
Jorabridge	5	0.13	3,323
Baishnabghat	5	0.59	5,916
Chingrighata	22	0.78	9,405
Pagladanga	30	1.13	13,160
Topsia	31	1.19	33,542
Ambedkar	31	1.19	34,517
Duttabagan	5	1.37	49,788
Cossipore	4	3.10	95,471
Ballygunge	20	5.71	225,074
Dhapa Lock	19	12.19	378,753
Palmer's Bridge	35	20.42	890,681
Total	**232**	**47.98**	**1,742,649**

Table 4. Summary statistics for estimated geographic area and population size of catchment areas upstream of sampled pumping stations for Kolkata.

	Minimum	Maximum	Average	Std. Dev.
Geographic Area (sq. km.)	0.06	20.42	3.69	6.04
Population Size	1,493	890,681	134,050	252,990

In total, 232 wastewater samples were collected from selected pumping stations for ES of typhoid and paratyphoid fever in Kolkata. More samples were typically collected at pumping stations that served larger catchment populations. The exceptions to this were the Kulia Tangra PS, which had a small catchment population (1,493 people) and a high number of samples collected from it (N = 20), and the Duttabagan and Cossipore facilities, which had relatively large catchment populations (49,788 and 95,471 people, respectively) but less samples collected from them (N = 5 and 4, respectively).

A list of the wards identified within each pumping station's upstream catchment area and the proportion of each ward's total area that lies within the catchment area is shown in Table 5. The number of wards captured within a PS catchment area ranged from 1 ward for the Bangur PS to 41 wards for the Palmer's Bridge PS. The values for the percentage of a ward's total area that was captured by sampled pumping station ranged from <1% (i.e., a small portion of the ward overlapped with the catchment area) to 100% (i.e., the entire ward was located within the catchment area). The information provided in Table 5 used in conjunction with Figure 10 (a map of the pumping station's upstream catchment areas) are pertinent for interpretation and actionability of the ES results.

Table 6 shows the distribution of wards by pumping station catchment area (i.e., how much each ward constitutes a sample that is collected from a pumping station). For example, a wastewater sample from the Kulia Tangra PS is 84% from Ward 57 and 16% from Ward 58. The range in values is similar to Table 5, with some wards making up <1% of a wastewater sample from a pumping station (see Ballygunge PS) and other wards constituting 100% of the sample (Ward 93 for Bangur PS).

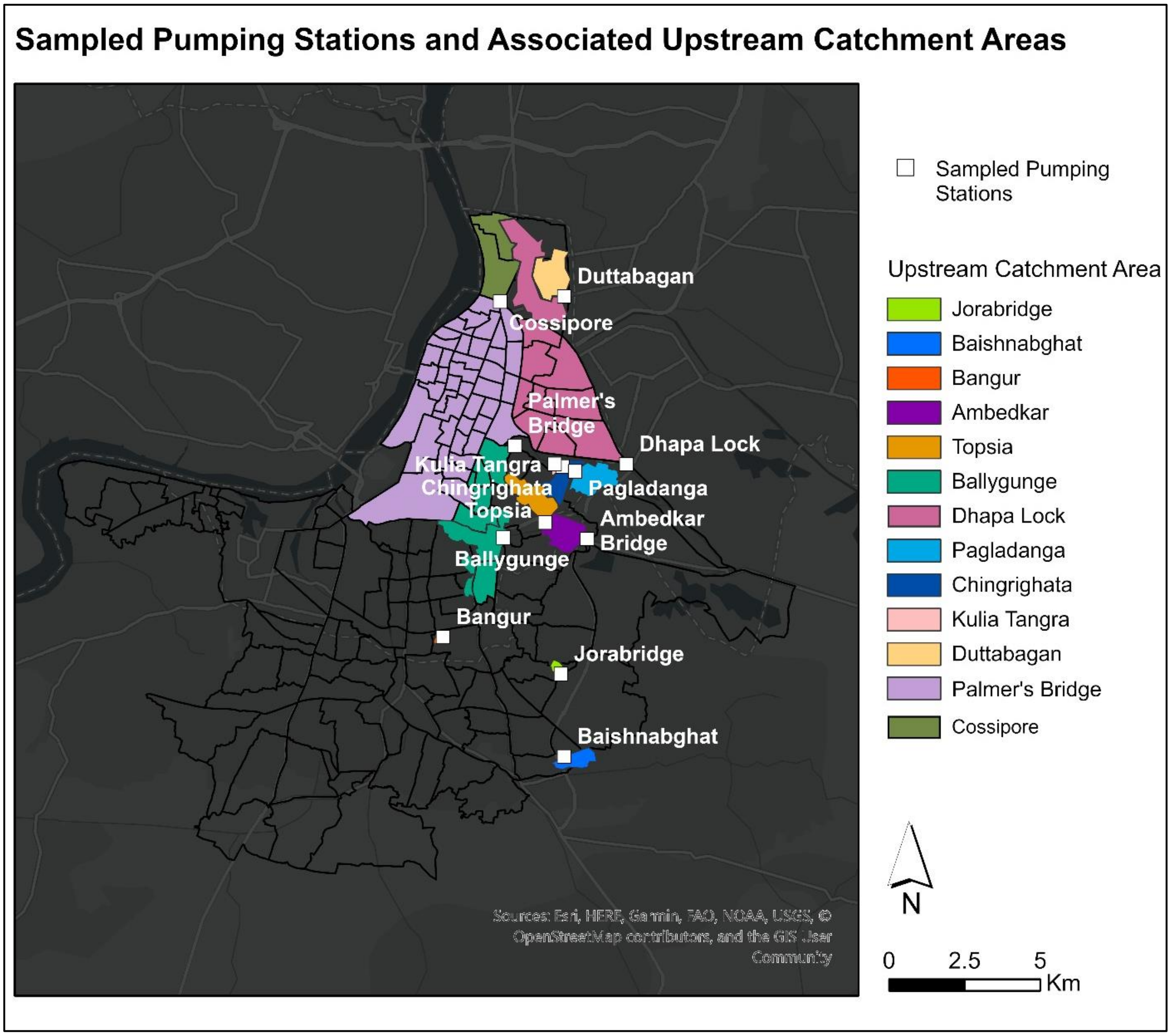


Figure 10. Mixed-methods, hybrid model used to map the upstream catchment areas of sampled pumping stations in Kolkata, India.

Table 5. Wards identified within upstream catchment areas of sampled pumping stations and the percentage of the total ward area that is contained within the catchment area.

Pumping Station	Wards Identified Within Catchment Area	Percentage of Total Ward Area Contained within PS Catchment Area
Kulia Tangra	57	3
	58	0.2
Bangur	93	3
Jorabridge	104	8
	106	2
Baishnabghat	109	0.4
	110	21
Chingrighata	57	6
	58	6
	66	1
Pagladanga	57	17
	58	5
Topsia	55	1
	56	18
	58	5
	59	21
	66	6
Ambedkar	58	0.1
	59	0.4
	66	35
Duttabagan	2	7
	3	45
	4	63
	5	2
Cossipore	1	100
	6	100
Ballygunge*	54 - 55, 60, 64, 68	100
	65	69
	69	33
	85	1
	86	20
Dhapa Lock*	13 - 14, 29 - 35	100
	5	55
Palmer's Bridge*	7 - 12, 15 - 28, 36 - 53, 61 - 63	100

*All or some of wards assumed to be contributing 100% of ward's area based on available sewer diagrams[44] and KMC source (2005) cited in Mukherjee (2020)[45]

Table 6. Distribution of wards by pumping station catchment area.

Pumping Station	Wards Identified Within PS Catchment Area	Percentage of PS Catchment Area by Ward
Kulia Tangra	57	84
	58	16
Bangur	93	100
Jorabridge	104	77
	106	23
Baishnabghat	109	7
	110	93
Chingrighata	57	31
	58	64
	66	5
Pagladanga	57	62
	58	38
Topsia	55	1
	56	10
	58	38
	59	34
	66	17
Ambedkar	58	0.5
	59	1
	66	99
Duttabagan	2	9
	3	45
	4	44
	5	2
Cossipore	1	44
	6	56
Ballygunge	54	7
	55	18
	60	10
	64	15
	65	17
	68	16
	69	12
	85	0.2
	86	3
Dhapa Lock	5	8
	13	8
	14	10
	29	5
	30	6
	31	14

	32	16
	33	18
	34	8
	35	9
Palmer's Bridge	7	2
	8	2
	9	1
	10	2
	11	2
	12	3
	15	3
	16	2
	17	2
	18	1
	19	1
	20	1
	21	2
	22	2
	23	1
	24	1
	25	2
	26	2
	27	2
	28	2
	36	5
	37	2
	38	2
	39	1
	40	2
	41	1
	42	2
	43	1
	44	3
	45	8
	46	6
	47	2
	48	1
	49	1
	50	2
	51	1
	52	1
	53	2
	61	3
	62	2
	63	19

Case Study #2: Accra, Ghana

Seven strategic ES sample collection locations were identified and are illustrated in Figure 11. Preference was given to pour points whose upstream catchment areas had a relatively higher population count, lower SES status, higher rate of open defecation, and/or history of cholera outbreaks[61] because these factors were assumed to potentially have a positive association with COVID-19 infection. The upstream catchment areas captured by these seven sampling points is shown in Figure 12. These seven locations were selected to illustrate how this method can be used to identify sampling locations with optimal population coverage without overlapping catchment populations. More locations could be added to achieve higher population coverage in the model, and also the locations of the sampling points could be toggled upstream or downstream to increase or decrease the catchment area size, respectively (depending on whether the user would like to have a larger or smaller population captured in the sample).

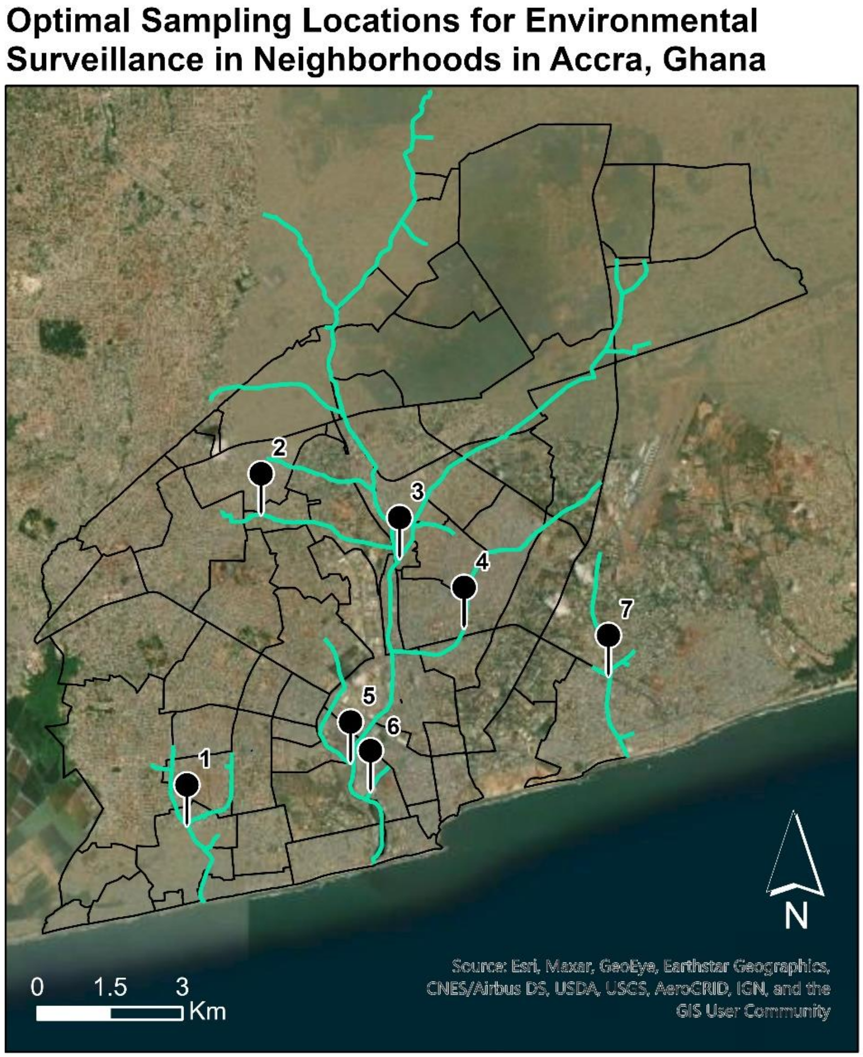


Figure 11. Seven strategic locations to collect samples for ES for SARS-CoV-2 in Accra, Ghana were identified and are shown here.

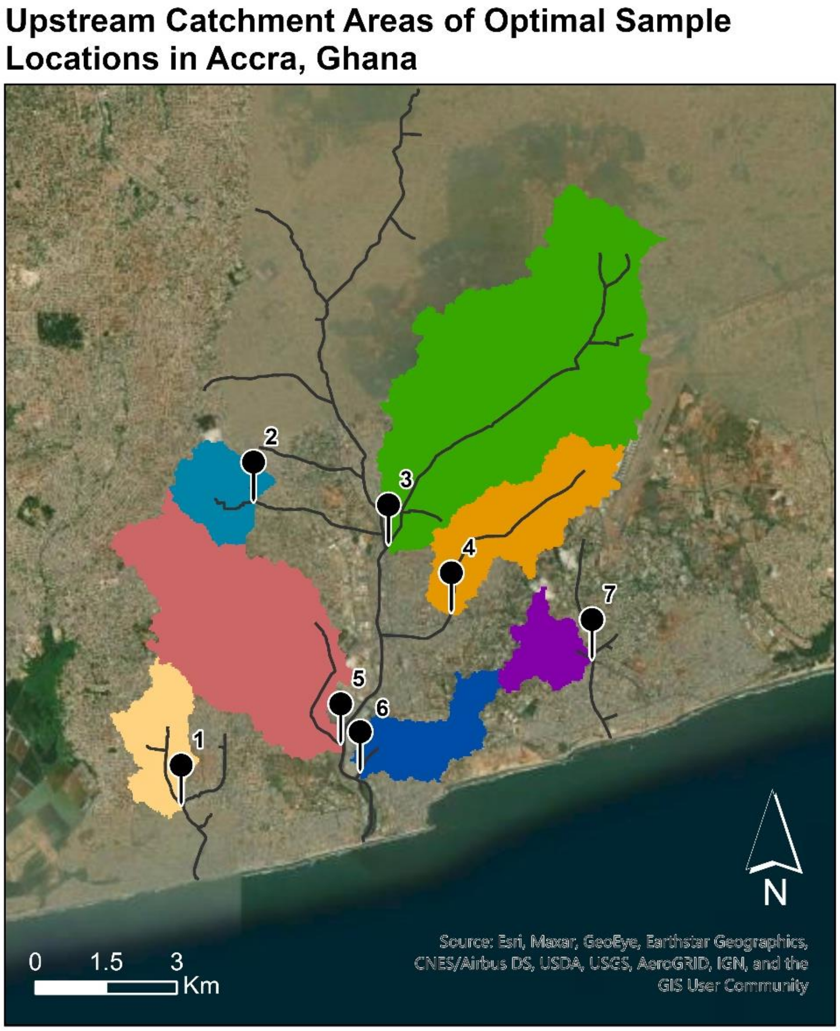


Figure 12. Upstream catchment areas of optimal locations to collect samples for ES for SARS-CoV-2 in Accra, Ghana.

The GPS locations of the proposed strategic sampling points and their upstream catchment size (geographic area and population size) are listed in Table 7. The geographic size of catchment areas ranged from 1.63 – 23.25 km^2, and the estimated population size ranged from approximately 17,032 – 253,611 people. The average catchment area was 7.86 km^2 and included approximately 109,244 people. The standard deviations for the geographic area and population size of the catchment areas were 7.99 km^2 and 81,647 people, respectively. These summary statistics are shown in Table 8. The Pearson correlation test indicated a positive linear relationship between geographic area and estimated population size of the catchment areas (r = 0.70).

Table 7. Proposed strategic sample locations and the estimated geographic area and population size captured by these locations in Accra.

Strategic Sampling Point	Latitude	Longitude	Geographic Area Served (sq km)	Population Size Estimate
1	5.5371562°N	0.2559009°W	3.62	61,885
2	5.5948968°N	0.2421168°W	3.04	78,410
3	5.5867519°N	0.2164054°W	23.25	153,280
4	5.5738117°N	0.2044346°W	6.01	152,001
5	5.5488370°N	0.2255134°W	14.30	253,611
6	5.5435215°N	0.2218098°W	3.21	48,489
7	5.5649079°N	0.1776306°W	1.63	17,032
Total	-	-	55.05	764,708

Table 8. Summary statistics for estimated geographic area and population size of catchment areas upstream of optimal sampling locations for Accra.

	Minimum	Maximum	Average	Std. Dev.
Geographic Area (sq km)	1.63	23.25	7.86	7.99
Population Size	17,032	253,611	109,244	81,647

A list of the neighborhoods identified within each upstream catchment area for each strategic sampling location is presented in Table 9, along with the proportion of each neighborhood's total area that is contained within the catchment area. The number of neighborhoods captured within the catchment areas of the proposed sampling points ranged from four neighborhoods (captured by Point 7) to 17 neighborhoods (captured by Point 3). The values for the percentage of a neighborhood's total area that was captured by a proposed point ranged from <1% to 100%. The information provided in Table 9, used in conjunction with Figure 12, (which shows upstream catchment areas of the strategic sampling points) are pertinent for interpretation and actionability of future ES results.

Table 9. Accra neighborhoods identified within upstream catchment areas of strategic sampling locations and the percentage of the total neighborhood area that is contained within the catchment area.

Strategic Sampling Point	Neighborhoods Within Catchment Area	Percentage of Total Neighborhood Area Contained within Catchment Area
1	Dansoman	29
	Mamponse	15
	New Mamprobi	10
	Russia	40
	South Odorkor	0.2
	Sukura	42
	West Abbossey Okai	15
2	Abeka	54
	Abeka Lapaz	5
	Bubuashie	5
	Darkuman	36
	New Fadama	23
	Nii Boi Town	6
3	Abelemkpe	51
	Abofu	1
	Achimota	45
	Airport Residential Area	64
	Alajo	62
	Anumle	18
	Dzorwulu	100
	East Legon	31
	Kotobabi	100
	Kpehe	84
	Legon	35
	Legon Staff Village	49
	Mamobi	11
	Mempeasem	8
	New Town	40
	Okponglo	100
	Roman Ridge	93
4	Airport Residential Area	36
	Kanda	47
	Kokomlemle	22
	Kotobabi	0.02
	Mamobi	89
	New Town	43
	Nima	73
	Roman Ridge	7

5	Abbossey Okai	100
	Bubuashie	85
	Dansoman	0.01
	Darkuman	56
	Kaneshie	84
	Korle-Bu	9
	Kwashieman	12
	Lartebiokorshie	23
	Mataheko	100
	North Kaneshie	15
	North Odorkor	15
	Russia	20
	Sabon Zongo	100
	South Odorkor	24
	West Abbossey Okai	85
	Zoti	81
6	Accra Central	63
	Adabraka	6
	Adedenkpo	38
	Korle Dudor	80
	Ridge	38
	Tudu	50
	Ussher Town	1
7	Kanda	2
	Osu	24
	Ridge	3
	Ringway	89

Table 10 shows the distribution of neighborhoods for each strategic sampling point's catchment area (e.g., a wastewater sample from Point 7 is made up of: 3% Kanda, 4% Ridge, 40% Osu, and 53% Ringway). These values ranged from <1% to 53%, indicating that there were no sampling points that collect wastewater from only one neighborhood.

Table 10. Distribution of neighborhoods for each strategic sampling point's catchment area.

Strategic Sampling Point	Neighborhoods Identified Within Catchment Area	Percentage of Strategic Sampling Point Catchment Area by Neighborhood
1	Dansoman	51
	Mamponse	7
	New Mamprobi	2
	Russia	13
	South Odorkor	0.2
	Sukura	16
	West Abbossey Okai	10
2	Abeka	45
	Abeka Lapaz	1
	Bubuashie	5
	Darkuman	43
	New Fadama	6
	Nii Boi Town	0.5
3	Abelemkpe	4
	Abofu	0.1
	Achimota	13
	Airport Residential Area	11
	Alajo	5
	Anumle	2
	Dzorwulu	16
	East Legon	8
	Kotobabi	6
	Kpehe	1
	Legon	17
	Legon Staff Village	4
	Mamobi	1
	Mempeasem	1
	New Town	3
	Okponglo	1
	Roman Ridge	7
4	Airport Residential Area	24
	Kanda	17
	Kokomlemle	6
	Mamobi	21
	New Town	11
	Nima	18
	Roman Ridge	2
5	Abbossey Okai	5
	Bubuashie	16
	Darkuman	14

	Kaneshie	16
	Korle-Bu	1
	Kwashieman	2
	Lartebiokorshie	3
	Mataheko	6
	North Kaneshie	2
	North Odorkor	2
	Russia	2
	Sabon Zongo	3
	South Odorkor	7
	West Abbossey Okai	14
	Zoti	7
6	Accra Central	18
	Adabraka	4
	Adedenkpo	13
	Korle Dudor	33
	Ridge	25
	Tudu	7
	Ussher Town	0.1
7	Kanda	3
	Osu	40
	Ridge	4
	Ringway	53

Comparison of Case Studies

The variation in geographic areas was similar between Kolkata (σ = 6.04) and Accra (σ = 7.99), whereas the variation in estimated population sizes captured by various sampling points was quite different between the two settings. There was considerably less variation in Accra (σ = 81,647) relative to Kolkata (σ = 252,990), which will be examined in the Discussion section. The Pearson correlation coefficient measuring the linear relationship between geographic area and population size indicated a stronger relationship for Kolkata (r = 0.99) than for Accra (r = 0.70). When the geographic coverage of the catchment areas was calculated for both settings (i.e., the sum of the catchment areas divided by the total area of the municipality) the geographic coverage rate was 42% for Accra and 24% for Kolkata. The calculated catchment population coverage (i.e., the total catchment area population divided by the total population in the municipality) was 46% for Accra and 39% for Kolkata.

Discussion

Case Study #1: Kolkata, India

The two main objectives of this thesis project were to: 1) address information gaps regarding the movement of human excreta throughout Kolkata and Accra to support the evidence generated by the ES project; and 2) develop a mixed-methods approach to build hybrid, context-specific catchment area models using free, open-access data and demonstrate its advantages when applied to ES. The development of the Kolkata model and the subsequent application and analysis led to valuable information on geographic location and estimated population size of catchment areas for selected pumping stations. The estimates catchment sizes were used to interpret ES data from previously collected wastewater samples that were analyzed for *S*. Typhi and *S*. Paratyphi A.

The catchment area model for Kolkata indicated that a wide range of population sizes were captured by the sampled pumping stations. These estimates were grouped into three categories: 1) <10,000 people; 2) 10,001 – 100,000 people; and 3) 100,001 – 1,000,000 people. The total number of wastewater samples collected for each category was 57, 101, and 74 (respectively). The large number of wastewater samples distributed across three orders of magnitude in population size will allow future analyses to:

> 1) Compare the concentration of human mitochondrial DNA markers detected in these wastewater samples to the estimated population size across the range of categories and determine if this marker can be used to "normalize" the pathogen PCR signal/concentration (*S*. Typhi and *S*. Paratyphi A) against an indicator of human excreta concentration and the size of the population contributing to that sample (i.e., higher concentrations of human mitochondrial DNA in wastewater suggest more concentrated human feces from a larger population);
>
> 2) Estimate the *S*. Typhi and *S*. Paratyphi A prevalence in different sizes of catchment populations based on the relationship between the PCR signal for these pathogens, the human mitochondrial DNA signal, and the estimated catchment population size. For Wards 58 and 59 in

Kolkata, the prevalence of *S*. Typhi and *S*. Paratyphi A infection based on wastewater detection will be compared to a community-based active surveillance system for these infections.

It was expected that pumping stations that served larger catchment sizes would have higher installed capacities and vice versa. The installed capacity figures were found for some of the sampled pumping stations on the Government of West Bengal's Irrigation and Waterways Department website page[43] (see Table 1). This trend was generally observed; however, it was noted that the Ballygunge PS, which has the largest installed capacity of the pumping stations in KMC (1,233 cusecs), had a smaller catchment geographic areas and population size than the Palmer's Bridge PS (1,184 cusecs) (see Table 3 for catchment area estimates). This finding may indicate that the estimates for these facilities have some uncertainty, and perhaps the assumptions that were made for the Ballygunge PS calculations resulted in underestimation of the catchment area size. There may also be other reasons for this observation, and one potential explanation is that the pumping stations are situated at different elevations. The elevation of the Ballygunge PS is at 3m, and the elevation of the Palmer's Bridge PS is at 7m[62]. The lower elevation of the Ballygunge PS may imply that pumps in this facility need more energy to oppose gravity to lift wastewater out of the facility, and this could explain why it needs a higher installed capacity than the Palmer's Bridge PS. Another potential explanation is that the geographic area the Ballygunge PS serves may experience more severe flooding than other areas, and therefore it may have a higher installed capacity to pump out floodwater during the monsoon season. These types of considerations can be investigated as next steps in an attempt to verify the catchment area estimates from this work.

The catchment area estimates derived from the Kolkata mixed-methods model were compared to Novel-T's estimates provided in their ES catalogue[28] and are shown in Table 11. Several of the pumping stations did not have estimates from Novel-T, so this project was able to contribute data for these sampled pumping stations to inform interpretation of ES results. Key differences between these estimates are that Novel-T estimated the watershed area draining to pumping stations and did not incorporate underground sewer and storm water lines. It also appears that DEM Reconditioning was not used in the Novel-T model. The mixed-methods model developed through this thesis project incorporated information about relevant hydraulic features and used DEM Reconditioning to improve the accuracy of the hydrologic model. Based on these differences, the catchment area estimates derived from the mixed-methods model are likely more accurate than Novel-T's watershed-based estimates. This assertion is supported by Novel-T's population size estimates for the Palmer's Bridge (16,087 people), Ballygunge (8,199 people), and Dhapa Lock (59,890 people) pumping stations, which are significantly smaller than the estimates derived from the mixed-methods model (890,681 people, 225,074 people, and 378,753 people, respectively). These three stations have the highest installed capacities in KMC (see Table 1 for reference) and are mentioned in the literature as facilities that serve large sewer and storm water networks[43,45]. Therefore, it is expected that these facilities would serve larger catchment areas similar to the numbers estimated by the mixed-methods model. The differences between the models for these three particular facilities showcases the value added by incorporating information about underground sewer and storm water lines that lead to these facilities. Conversely, Novel-T's estimates for the Chingrighata PS were significantly larger than the mixed-methods model estimated (Novel-T: 18.87 km^2 and 892,646 people; Kolkata Mixed-Methods Model: 0.78 km^2 and 9,405 people). It is highly unlikely that the Chingrighata PS could sustain the volume of wastewater to support Novel-T's estimates with its installed capacity of only 100 cusecs[43]. In addition, Novel-T's catchment population size estimate for the Kulia Tangra PS was only 57 people, whereas the mixed-methods model estimated 1,493 people. Novel-T's low estimate for this facility is unlikely due to the high population density throughout Kolkata and would imply that this facility serves a very small geographic area. These findings demonstrate the value that incorporating information about sanitation infrastructure adds to catchment area estimates when a mixed-methods approach is employed. In addition, the DEM Reconditioning tool has been shown to improve the accuracy of free DEMs in watershed modeling, and therefore its use increased confidence in the estimates derived from the mixed-methods model.

Table 11. Comparison of estimated catchment sizes (geographic area and population) of sampled pumping stations for ES in Kolkata between Novel-T and the mixed-methods model developed through this thesis project.

	Geographic area (sq-km)		Estimated Population Size	
Pumping Station	**Novel-T**	**Mixed-Methods Model**	**Novel-T**	**Mixed-Methods Model**
Ambedkar	<1	1.19	2,262	34,517
Baishnabghat	NA	0.59	NA	5,916
Ballygunge	<1	5.71	8,199	225,074
Bangur	NA	0.06	NA	1,525
Chingrighata	18.87	0.78	892,646	9,405
Cossipore	NA	3.10	NA	95,471
Dhapa Lock	2.07	12.19	59,890	378,753
Duttabagan	NA	1.37	NA	49,788
Jorabridge	NA	0.13	NA	3,323
Kulia Tangra	<1	0.12	57	1,493
Pagladanga	<1	1.13	7,395	13,160
Palmer's Bridge	<1	20.42	16,087	890,681
Topsia	<1	1.19	3,232	33,542

NA = Not Available

The impact of DEM Reconditioning in the Kolkata model is evident when the results of the Flow Accumulation tool are compared for a raw DEM versus a reconditioned DEM. As Figure 13 illustrates, using the raw DEM as the input results in several instances where the canals and streams identified in OSM and Google Maps are not identified by the Flow Accumulation tool. It is therefore not a very accurate reflection of the actual drainage system. While this can be attributed to the flat terrain of Kolkata and low-resolution quality of open-access DEMs, the results suggest that the natural drainage patterns in Kolkata do not coincide with its constructed network of canals. Due to its outdated sewerage system that was designed in the 1800s, coupled with rapid urban expansion, Kolkata’s canals may not have been constructed along optimal drainage routes. If open canals were constructed in areas with high flow accumulation, it may help drain low-lying areas that routinely accumulate standing water in Kolkata.

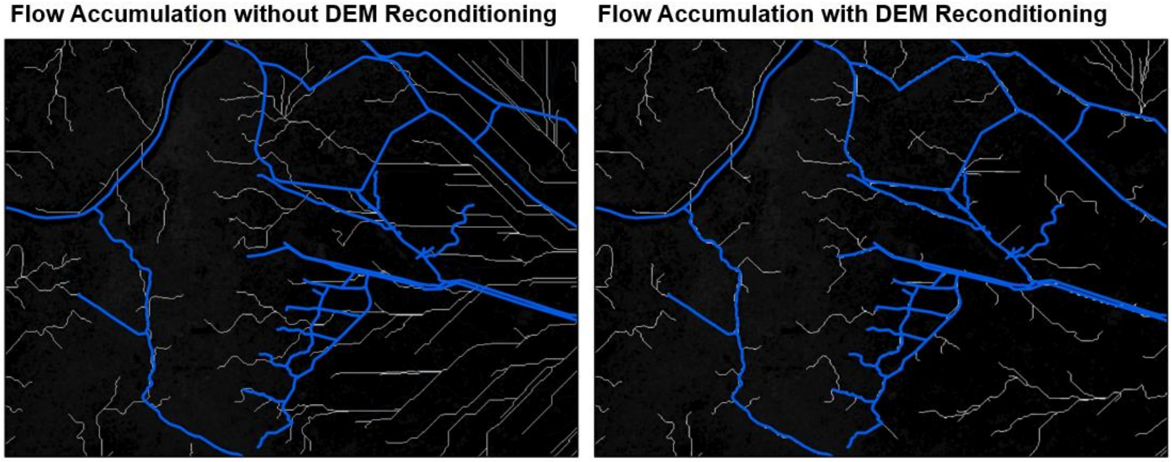


Figure 13. Comparing the drainage network calculated by the Flow Accumulation tool on a DEM without (left) and with (right) DEM Reconditioning. The outputs demonstrate that the DEM Reconditioning tool results in a drainage network that is a more accurate reflection of the true network of waterways.

Although the results of DEM Reconditioning appear promising because the drainage network more closely resembles the actual waterways, it is possible that burning streams into the DEM for Kolkata may skew the results. The flow accumulation network generated from the unmanipulated DEM creates a stream network that primarily flows in an easterly direction away from the Hooghly River, which is consistent with the literature on Kolkata's general direction of drainage. After DEM Reconditioning, the stream network flows towards the burned-in waterways, but these forced drainage patterns may be over-emphasized. In other words, it is possible that due to the lack of surface variation in Kolkata, burning-in waterways may have too strong of an impact on the hydrologic model. The attempt made to control for this was to enter significantly smaller values for the "Sharp drop in Z units" field than the default settings[2] and to burn waterways based on their relative size classification. The intention behind these control measures was to reduce the weight of the burned-in stream network to create a more realistic hydrologic

[2] Default settings in DEM Reconditioning tool: number of cells for stream buffer = 5, smooth drop in Z units = 10, and Sharp drop in Z units = 1000 [26]

network for Kolkata. Despite these efforts, the outputs from the Flow Accumulation tool before and after DEM Reconditioning may indicate that DEM Reconditioning for settings with flat terrain and extensive artificial drainage networks bias the drainage patterns away from the truth.

The total geographic area and population size captured by the catchment areas resulted in a coverage rate of 24% and 39% (respectively) in the KMC. There was a wide range of values in the geographic area and population size estimates, which is a meaningful finding because it affects how sample results are interpreted. For instance, samples taken from pumping stations with smaller catchment populations, such as the Kulia Tangra and Bangur facilities, may not be ideal for predicting disease burden at the city-level. Results from these sample sites are much more actionable, however, because there is a smaller catchment area to narrow down where cases of typhoid fever may be located. Samples taken from pumping stations with large, highly populated catchment areas are useful for estimating the local disease burden at the time of the sample, but an aggressive response will require more resources to identify locations of typhoid fever cases. This is particularly true for samples collected from the Palmer's Bridge PS in which the geographic area and population size of the catchment area are 2.77 and 2.99 standard deviations above the average, respectively. If Palmer's Bridge is removed from the calculations, the standard deviation of the geographic catchment area reduces from 6.04 to 3.49 km^2, and for the catchment population size it reduces from 252,990 to 115,934 people. This finding has important implications for how results from samples collected from this facility are interpreted and compared to results from samples from other pumping stations. Due to the presence of significant volumes of wastewater from other sources (e.g., from bathing, washing, industrial processes, etc.) and the substantial distance pathogens must travel to reach this pumping station, the results from these samples may indicate a low concentration of *S*. Typhi or *S*. Paratyphi A in wastewater that is not an accurate reflection of the upstream disease burden in the catchment population. This is particularly problematic after rain events and during the monsoon season because rainwater runoff will be included in the sample and cause further dilution.

Rainfall has a dynamic, profound impact on the content of samples taken at pumping stations and their corresponding catchment areas. This is true for any urban setting where ES is implemented, but it is particularly important to consider in the context of Kolkata due to its combined sewer and storm water system and insufficient pumping capacity. This is most clearly demonstrated by the Palmer's Bridge PS and Ballygunge PS. During dry periods, these facilities route wastewater through high-level sewers to the Topsia Point 'A' PS. During and after heavy rain events, effluent from these facilities is discharged directly into SWF channels due to the high-level sewer's inability to cope with the increased volume[45]. There is uncertainty about whether the Topsia Point 'A' PS was sampled during the recent ES activities because there are two pumping stations with similar names (Topsia and Topsia Point 'A'), and it is unclear from data collection surveys whether these are actually referring to the same facility. This information has implications for the interpretation of the lab results because the catchment population at Topsia Point 'A' changes when it rains. A next step to resolve this uncertainty will be to contact partners in Kolkata to clarify which pumping station was sampled.

All of the KMC pumping stations naturally experience shifts in their upstream catchment areas to some extent when it rains. This is due to the nature of the combined sewage and storm water drainage system and its means of conveyance through open canals. If the storm and sewerage systems were separated and if sewage were contained within closed pipes, rainfall would not add to the volume of wastewater generated by residents in KMC and increase the total volume in the drainage system. However, because rainfall cohabits the drainage system, wastewater at pumping stations contains rainwater runoff (and the contaminants it seizes en route to the facility) in addition to its regular DWF. This may have two opposing effects that have implications for interpretation of the lab results: 1) the rainfall dilutes blackwater in the sewerage system resulting in a lower concentration of pathogens in the wastewater samples; 2) rainwater runoff may increase the concentration of pathogens in the wastewater samples due to human excreta in the environment that gets washed into the sewerage system.

The interwoven network of closed, underground sewer and storm water lines combined with open canals was very difficult to map, and many assumptions were made due to gaps in the available information. Although the mixed-methods model for Kolkata was able to incorporate relevant information about sanitation infrastructure and synthesize it with hydrologic processes, there were still some significant unknowns where additional information would have improved the accuracy of the catchment size estimates. Sewerage diagrams were not available for underground sewer and storm water lines leading to pumping stations other than the Palmer's Bridge, Ballygunge, and Dhapa Lock facilities. Due to this, it was assumed that underground sewer and storm water lines would generally follow the direction of flow calculated by the hydrology tools in ArcGIS. This assumption was based on the knowledge that gravity will be leveraged as much as possible for wastewater conveyance to reduce energy usage and costs. This is an acknowledged limitation of the Kolkata model, however, and leads to the presumption that the catchment population size estimates for pumping stations without underground connection information are underestimated. In addition, the use of 2011 census data further contributes to the likelihood that our population size estimates will be lower than the true values. These sources of uncertainty and error should be reviewed with local partners to fill knowledge gaps and approximate the magnitude of error of the current catchment population size estimates.

Case Study #2: Accra, Ghana

The development of the Accra model and the subsequent analysis demonstrated how a mixed-methods approach that uses free, open-access DEMs can be used to identify strategic sampling points along waterways to collect samples for ES of COVID-19 in the future. The development of this second model demonstrated how this process can be adapted and tailored for different settings. In addition, the findings show the advantages of applying this methodology prior to sample collection to ensure adequate coverage of the population(s) of interest is achieved.

The catchment area model for Accra found that a wide range of geographic areas and population sizes can be captured by the proposed sampling points along major waterways. It is important to highlight that the catchment sizes can be adjusted by moving the sampling points either upstream or downstream to decrease or increase the upstream catchment size, respectively. These points were proposed because they capture a representative range of geographic locations, land area, and population sizes. The outputs can also be used to demonstrate the utility of this methodology for ES sample site selection. The methods and results from this project are currently being shared with local partners in the TREND Group in Accra to use as they move forward on planning and operationalizing a new COVID-19 ES system. These conversations will enable this research project to transition from theoretical demonstrations to sample collection at the recommended locations.

The Accra model developed for this thesis project was compared to Novel-T's model for Accra located in their ES catalogue of Environmental Sites[28]. Upon inspection, it appeared that the drainage lines in the Novel-T map closely mirrored the actual waterways, suggesting that DEM Reconditioning was applied for this site. There was only one watershed in the Novel-T model that was suitable for comparison with the results from our proposed Accra model. Novel-T's Nima Freetown watershed most closely resembled Point 4 in the proposed model for Accra; however, its population size was significantly less (77,019 people) compared to the Accra model developed herein (152,001 people). A possible explanation for this discrepancy stems from Novel-T's use of WorldPop data for population estimates, while the proposed Accra model relied on data from the 2010 census. Although the census data is outdated, previous comparisons of these two data sources for this project showed that WorldPop may significantly underestimate the true population size in this area. This could be due to a high number of people living in high-rise buildings that are not captured in the WorldPop data, which is calculated from satellite imagery. Although the use of census data from 2010 is a limitation of the proposed Accra model, there is reason to believe it leads to more accurate population estimates than WorldPop-derived estimates in dense urban areas with many high-rise apartment buildings. The other three sites in the Novel-T map did not resemble the proposed sampling locations in the Accra model closely enough to be directly compared because they were not placed at pour points in the waterways that capture large upstream areas. Novel-T's other sites captured relatively small catchment population sizes (< 2,833 people), and it is unclear why these sites were selected for analysis by Novel-T. The proposed model for strategic sampling points in Accra therefore provides new information about potential sites to collect samples for ES that capture human excreta from large, representative catchment areas.

DEM Reconditioning was very effective for Accra. As Figure 14 depicts, this tool adjusted the flow accumulation paths calculated in ArcGIS to match the presence of the true waterways. The adjustments were less drastic than were made to the Kolkata DEM because the flow accumulation paths calculated by ArcGIS were already very similar to the true waterways feature. This can be attributed to the hilly terrain of Accra because gravity is driving the flow of water and wastewater through the environment, and pumping stations are not present to oppose this force.

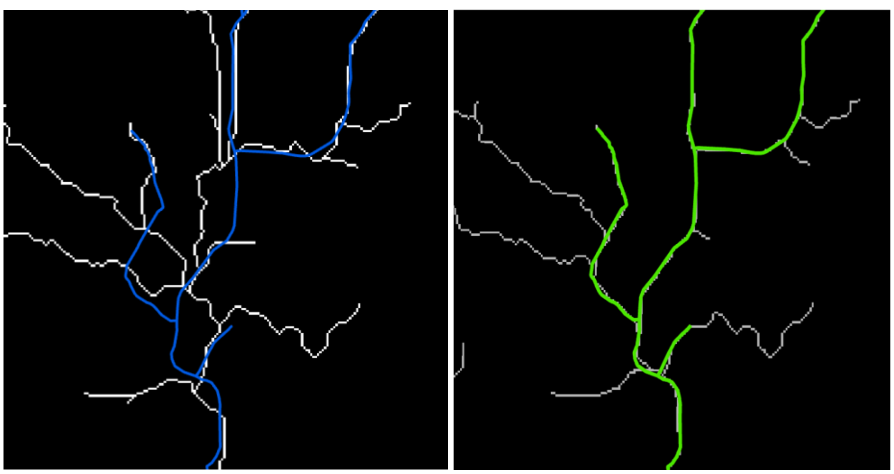


Figure 14. Before and after comparison for DEM Reconditioning step.

Cumulatively, the catchment areas of the proposed sampling locations resulted in 42% coverage of the total area and 46% coverage of the total population in the AMA. Importantly, there were multiple catchment areas along the coast that could not be captured by proposed sampling locations through this method because they drained towards the ocean and not directly into the waterways. This finding has important implications because coastal communities may not be sufficiently represented by samples collected from the waterways in Accra. To capture representative samples of excreted pathogens in these neighborhoods, alternative sample locations and strategies (e.g., samples from pit latrines or septic tanks) may be required to ensure adequate coverage.

Rain events will have substantial effects on the catchment areas draining towards pour points in Accra due to the system of open drainage and continued practice of open defecation. After heavy rain events, feces on land or contained in open drains may be seized by runoff and drain to pour points. These sources will not be captured during dry conditions, so it will be important to monitor rain events in relation to sampling days.

There are several limitations of the Accra model that must be acknowledged. Although the model is able to capture excreta that is discharged into open drains by identifying downstream points that receive this wastewater, the accuracy and actionability of the model is still limited by the lack of comprehensive maps that show open drain connections for all of the AMA. Adding more information to the model will always improve its accuracy, and therefore if open drain maps are found for all the neighborhoods, these routes could be incorporated to improve catchment size estimates. Another source of uncertainty is the large volume of sewage that is vacuumed out of septic tanks and transported offsite for treatment/disposal in Accra. The proposed sampling points will not capture excreta from individuals who defecate in onsite sanitation facilities from which sewage and/or fecal sludge is disposed of in this manner. This uncertainty implies that catchment population size may be overestimated because these points capture all of the population that live upstream (according to 2010 census data). This source of error is likely offset by the use of outdated census data to inform the population size estimates because it is probable that significant population growth has occurred since this census data was collected. These considerations should be discussed with local partners to gain a better understanding about the level of uncertainty and approximate magnitude and direction of error in the catchment population size estimates.

Comparison of Case Studies

The DEM Reconditioning tool created superior results for the Accra model compared to the Kolkata model. This is attributed to the differences in terrain and sanitation infrastructure between the two settings. Kolkata is situated in a mostly flat, slightly bowl-shaped landscape and heavily relies on hydraulic structures to drain and convey wastewater, whereas Accra is situated in a hilly environment and relies on gravity to convey wastewater through open drains into larger waterways. The terrain affects the ability of the hydrology tools in ArcGIS to capture true differences in elevation between pixels, especially when free DEM data is used due to the large pixel sizes. When there is greater variation in elevation values, such as in the case of Accra, these tools perform optimally. This resulted in improved accuracy in the flow accumulation outputs and, ultimately, the delineated catchment areas for the Accra model.

The Pearson correlation coefficient between geographic area and population size of the catchment areas both indicated a positive linear relationship, but there was a stronger correlation for Kolkata ($r = 0.99$) than for Accra ($r = 0.70$). This observation is logical and can be explained by a number of factors. The proposed sampling locations for Accra will capture areas upstream of their locations (see Figure 12), and these upstream areas may be outside of the AMA boundary where population density is lower. For Kolkata, however, the sampled pumping stations serve densely populated wards and are not able to capture large areas upstream of their locations due to the presence of the Hooghly River, which acts as a barrier between upstream runoff and the city of Kolkata.

The Accra model achieved significantly higher coverage rates for geographic extent and population size of its combined catchment areas. This was despite having nearly half the sample sites of Kolkata and excluding catchment areas that drained directly to the ocean. The most likely explanation for these differences is the relatively simplistic nature of Accra's drainage system compared to Kolkata's complex system. Accra's drainage system relies on gravity to convey wastewater through open drains and channels towards the ocean, and consequently, larger areas are more easily captured because sample points simply need to be moved further downstream. This is not so easily executed for Kolkata because its infrastructure is much more complex and difficult to trace.

There are limitations in both models that stem from known and unknown uncertainties. Known sources of error that contribute to potential error in the catchment population size estimates are the use of outdated census data to inform these population estimates and gaps in information about sanitation infrastructure in both settings. It is difficult to approximate the magnitude of error that arises from these sources of uncertainty, but this approximation could be informed by conversations with local partners to gain better understanding about the population dynamics and sanitation landscape in both settings. In addition, a suggested next step to try to quantify the error associated with this approach is to apply it to Dhaka, Bangladesh to estimate catchment population sizes and compare these numbers to Novel-T's results. Novel-T has a relatively high confidence rating (3 out of 5 stars) in their Dhaka estimates due to use of high resolution DEMs and "blue-lining" to confirm the location of waterways in the field[28].

Conclusions

There is growing recognition of the importance of ES as a public health tool, and its use is currently being tested in many cities around the world in response to the COVID-19 pandemic. Modeling catchment areas is a highly useful practice to improve ES for excreted pathogens in varying settings. The hybrid, mixed-methods approach described in this thesis is able to capture and visualize the dynamic processes that influence movement of human feces throughout the environment in different settings. This methodology synthesizes available information on sanitation infrastructure, demographics, and drainage patterns, and it integrates these elements into a hybrid model. By doing so, a city's unique sanitation landscape can be better represented, which is inherently relevant to designing efficient ES strategies and interpreting the results from the ES samples. Increasing knowledge about the drainage routes of wastewater is fundamental to interpreting the results of ES samples because it provides information about the catchment population contributing to the sample, such as its size, location, and key characteristics. When the estimates from the Kolkata and Accra models were compared to Novel-T's watershed-based estimates for these sites, it was shown that the methodology proposed herein likely led to more accurate catchment size estimates.

In the context of Kolkata, estimating the location, size, and population of catchment areas is valuable because these figures can be linked to ES lab results and clinical cases to estimate the burden of disease. This information is also important for enhancing knowledge about transport, fate, and detection limits of *S*. Typhi and *S*. Paratyphi A in uncontrolled, real-world settings. The knowledge gained from this process also increases the actionability of sample results by narrowing down likely geographic areas where there may be higher prevalence of disease. This is ideal when resources (i.e., funds, testing capacity, treatment, trained personnel, etc.) are limited, and it enables response teams to target more aggressive interventions to infection hotspots. Lastly, by cross-referencing catchment areas with relevant contextual data, the outputs enable researchers to investigate whether certain variables or exposures in the source populations are associated with frequency and/or magnitude of pathogen detection. Examples of census data that may be relevant to transmission of typhoid fever include type of water supply, percent of houses in poor condition, percent of houses made of temporary materials, and percent of houses with no latrines[42]. The variable "percent of houses with no latrines" is shown in Figure 15 to illustrate how mapping this information can be used in conjunction with the Kolkata catchment area map (Figure 10) to link potentially relevant factors to lab results.

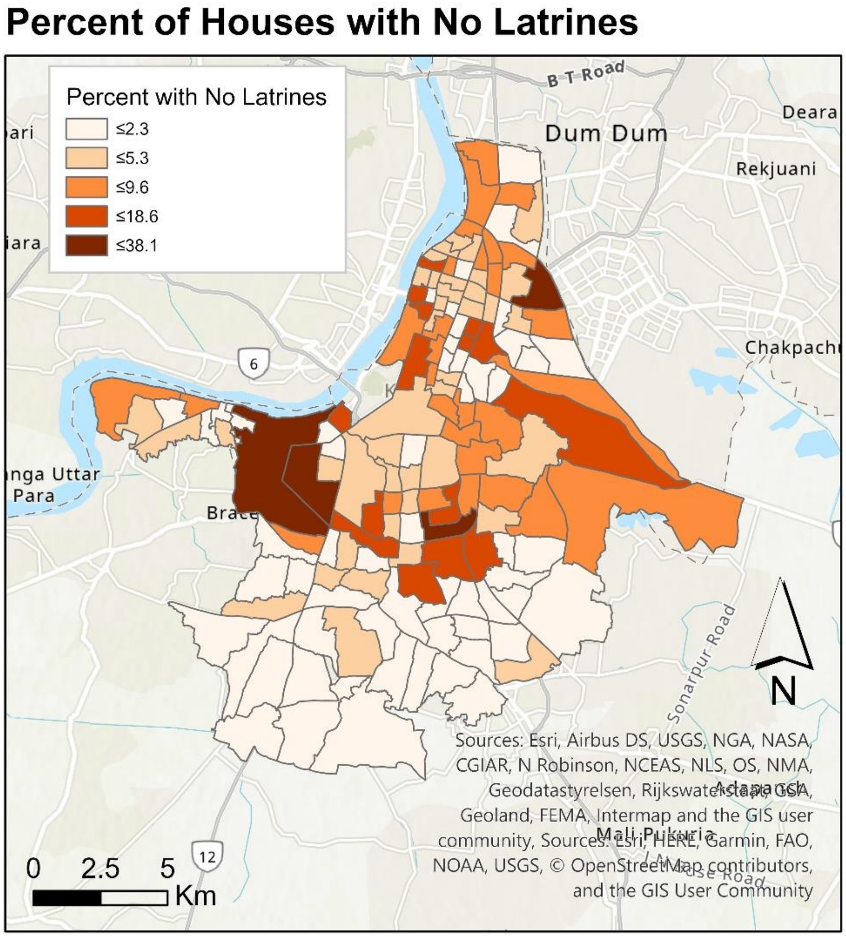


Figure 15. Percent of houses with no latrines in KMC wards according to 2011 Census of India.

Applying this methodology to Accra during the design and planning stages of COVID-19 ES has the added benefit of allowing public health authorities to proactively select desired population sizes and prioritize subpopulations for ES activities. It is ideal to involve key stakeholders during this process to select sample points that capture the neighborhoods and/or characteristics of interest. For Accra, examples of possible criteria for selecting sample sites included population counts, low SES, history of cholera outbreaks, and open defecation rates. The results of the model will be shared with local health authorities to illustrate its potential. From there, the sample points can be adjusted to capture specific areas according to the priorities of the Accra stakeholders. If this methodology is implemented before ES activities have concluded, sample points can be adjusted throughout the length of the project. For instance, by simply toggling a sample point further upstream or downstream, the surveillance implementers can control how large or small the contributing catchment population is. This creates space for an ongoing dialogue with local stakeholders and allows flexibility in the sample site selection process. This is an important feature because priorities often shift throughout the life of a project, such as if an outbreak were to occur and specific communities needed to be more closely monitored.

Analyzing Kolkata and Accra as two separate case studies served a twofold purpose: it allowed the findings from this project to inform two real-world ES projects, and it explored the applicability of this methodology to two settings. These two cities differ greatly in their terrain and sanitation infrastructure. While this is a strength of the study because it explores the utility of catchment area modeling to support ES in different environments, it also makes it difficult to compare results between the two models.

The results from this analysis indicate that using free global DEM datasets for flat terrain resulted in less credible results because the resolution was too low to detect differences in elevation between the pixels. While this is a widely acknowledged limitation, a potentially new finding that arose from this research is that DEM Reconditioning may not be appropriate for areas with flat terrain and extensive artificial canals. This is because this tool may actually over-correct the hydrology network and create drainage patterns that do not actually exist. There may be more advanced methods that allow for sensitive manipulation of DEMs for flat terrain that could be used to address this concern, and alternatives to using the DEM Reconditioning tool for flat terrain with artificial drainage structures should be further explored in settings similar to Kolkata. Specifically, a suggested next step is to apply this methodology to Dhaka and compare the outputs to Novel-T's ES maps. These maps were derived from high-resolution DEMs and field surveys to confirm the location, size, and flow direction of waterways (a process coined "blue-lining"), lending to higher confidence in the results[28]. The availability of this information coupled with Dhaka's similarities to Kolkata in sanitation infrastructure, flat terrain, and population density make it an excellent setting to test the appropriateness of DEM Reconditioning on a 30-m DEM.

The findings from the catchment area model for Accra importantly illustrated that wastewater in neighborhoods along the coast will not all be captured by sampling points along the primary waterways. This finding may be applicable to other coastal communities, which is another reason why it is advantageous to implement this methodology prior to sample collection. The catchment area model may also identify zones within the city that are located too far away from sample points to be detectable due to pathogen die-off and excessive dilution. Gaining this knowledge allows public health authorities to design a sampling strategy that allocates samples appropriately to achieve equitable representation of the study area, such as collection of pit latrine samples in areas where excreta is not discharged into open drains and waterways.

There are several limitations of this methodology that need to be acknowledged. A major underlying assumption of the model for Kolkata is that, in areas where no underground sewerage information is available, the delineated catchment areas and flow accumulation network will serve as a drainage compass. This approach assumes that the flow in underground lines is in accordance with the gravity-fed direction of drainage and assumes there are no walls or structures that disrupt these patterns, so the outputs must be interpreted with caution. Additionally, it is not possible to assume that human movement in and out of catchment areas is negligible. For example, wastewater from commercial parts of the city may include excreted pathogens from people who live in different areas of the city. Therefore, the outputs from this tool may not always reflect actual cases of infection within a catchment area when the result from a wastewater sample is positive for the target pathogen, or that there are no cases in a catchment area when a result is negative. For this reason, the catchment areas should be referred to as "likely" or "less likely" to have cases of infection depending on the lab results from the wastewater samples. Furthermore, the models are not able to capture excreta that is stored within latrines or septic tanks and transported to off-site facilities. These containers are often emptied manually or with vacuum trucks, and the fecal sludge is disposed of at landfills or treatment plants. In some communities, latrines and septic tanks may directly discharge into open drains and canals, but this data would need to be collected for each sample location. A future recommendation for all ES programs is to record this information during sample collection.

Further limitations arise from the development of the model itself. All mapping operations were performed manually in ArcGIS Pro and are therefore innately subject to human error. To address this concern, a potential next step is to build this methodology by code. In theory, a raw DEM could be input into the model, and the sequential operations in ArcGIS could be executed without requiring human involvement beyond the initial set-up. It would be necessary to create models that take into consideration certain characteristics of the setting for which the model is being built (e.g., open drains versus closed sewer pipes, gravity-fed versus mechanically-driven systems) as these elements affect the movement of sewage in urban environments. Additionally, ArcGIS Pro is not free and requires significant funds to install the software and sustain annual subscriptions. The next step to address this limitation is to attempt to develop these models using QGIS instead of ArcGIS because it is free to use and install.

The development of the models for Accra and Kolkata was not influenced directly by, or created in collaboration, with local stakeholders. This was due to lack of accessibility to contacts and resources during the time frame when this thesis project was developed. This is a limitation of the model and the basis for a recommended next step that the findings from these models should be shared with local collaborators and experts in the near future.

The most recent census data for Accra and Kolkata are over a decade old, and therefore the population size estimates are outdated. As a next step, a growth rate can be applied to adjust the population estimates based on the city's estimated population growth since the time of the last census. If possible, local stakeholders should be consulted during this process as they may be able to provide more specific information about growth rates at the ward and neighborhood levels.

Finally, spatial and spatio-temporal analyses should be conducted to quantify relationships between catchment areas with pathogen detection and/or disease incidence. Spatial regression modeling could also be performed to investigate whether certain covariates are significant in Accra and Kolkata catchment areas.

REFERENCES

1. Polgreen PM, Polgreen EL. Emerging and Re-emerging Pathogens and Diseases, and Health Consequences of a Changing Climate. In: *Infectious Diseases*. Elsevier; 2017:40-48.e2. doi:10.1016/b978-0-7020-6285-8.00004-6

2. Thacker SB, Berkelman RL. Public Health Surveillance in the United States. *Epidemiol Rev*. 1988;10(1):164-190. doi:10.1093/oxfordjournals.epirev.a036021

3. Murray J, Cohen AL. Infectious Disease Surveillance. In: *International Encyclopedia of Public Health*. Elsevier Inc.; 2016:222-229. doi:10.1016/B978-0-12-803678-5.00517-8

4. Holmes KK, Bertozzi S, Bloom BR, et al. Major Infectious Diseases: Key Messages from Disease Control Priorities, Third Edition. In: *Disease Control Priorities, Third Edition (Volume 6): Major Infectious Diseases*. The World Bank; 2017:1-27. doi:10.1596/978-1-4648-0524-0_ch1

5. Sims N, Kasprzyk-Hordern B. Future perspectives of wastewater-based epidemiology: Monitoring infectious disease spread and resistance to the community level. *Environ Int*. 2020;139:105689. doi:10.1016/j.envint.2020.105689

6. Hovi T, Shulman LM, Van Der Avoort H, Deshpande J, Roivainen M, De Gourville EM. Role of environmental poliovirus surveillance in global polio eradication and beyond. *Epidemiol Infect*. 2012;140(1):1-13. doi:10.1017/S095026881000316X

7. Wang Y, Moe CL, Dutta S, et al. Designing a typhoid environmental surveillance study: A simulation model for optimum sampling site allocation. *Epidemics*. 2020;31. doi:10.1016/j.epidem.2020.100391

8. Bivins A, North D, Ahmad A, et al. Wastewater-Based Epidemiology: Global Collaborative to Maximize Contributions in the Fight against COVID-19. *Environ Sci Technol*. 2020;54(13):7754-7757. doi:10.1021/acs.est.0c02388

9. Chen P, Liu Y, Wang H, et al. Environmental surveillance complements case-based surveillance of acute flaccid paralysis in polio endgame strategy 2019-2023. *Appl Environ Microbiol*. 2020;86(15). doi:10.1128/AEM.00702-20

10. World Health Organization. *World Health Organization WHO Guidelines for Environmental Surveillance of Poliovirus Circulation Vaccines and Biologicals*.; 2003. www.who.int/vaccines-documents/. Accessed December 7, 2020.

11. Wang Y, Moe CL, Teunis PFM. Children Are Exposed to Fecal Contamination via Multiple Interconnected Pathways: A Network Model for Exposure Assessment. *Risk Anal*. 2018;38(11):2478-2496. doi:10.1111/risa.13146

12. Andrews JR, Yu AT, Saha S, et al. Environmental Surveillance as a Tool for Identifying High-risk Settings for Typhoid Transmission. *Clin Infect Dis*. 2020;71(Supplement_2):S71-S78. doi:10.1093/cid/ciaa513

13. Daughton CG. Wastewater surveillance for population-wide Covid-19: The present and future. *Sci Total Environ*. 2020;736:139631. doi:10.1016/j.scitotenv.2020.139631

14. Debes AK, Ateudjieu J, Guenou E, et al. Clinical and environmental surveillance for vibrio cholerae in resource constrained areas: Application during a 1-year surveillance in the Far North Region of Cameroon. *Am J Trop Med Hyg*. 2016;94(3):537-543. doi:10.4269/ajtmh.15-0496

15. Takane M, Yabe S, Tateshita Y, et al. Satellite imagery technology in public health: Analysis of site catchment areas for assessment of poliovirus circulation in Nigeria and Niger. *Geospat Health*. 2016;11(3):355-361. doi:10.4081/gh.2016.462

16. Hawker L, Bates P, Neal J, Rougier J. Perspectives on Digital Elevation Model (DEM) Simulation for Flood Modeling in the Absence of a High-Accuracy Open Access Global DEM. *Front Earth Sci*. 2018;6:233. doi:10.3389/feart.2018.00233

17. Tavares da Costa R, Mazzoli P, Bagli S. Limitations Posed by Free DEMs in Watershed Studies: The Case of River Tanaro in Italy. *Front Earth Sci*. 2019;7:141. doi:10.3389/feart.2019.00141

18. Courty LG, Soriano-Monzalvo JC, Pedrozo-Acuña A. Evaluation of open-access global digital elevation models (AW3D30, SRTM, and ASTER) for flood modelling purposes. *J Flood Risk Manag*. 2019;12(S1). doi:10.1111/jfr3.12550

19. Farr TG, Rosen PA, Caro E, et al. The shuttle radar topography mission. *Rev Geophys*. 2007;45(2). doi:10.1029/2005RG000183

20. SRTM - eoPortal Directory - Satellite Missions. https://earth.esa.int/web/eoportal/satellite-missions/s/srtm. Accessed January 26, 2021.

21. Khal M, Algouti A, Algouti A, Akdim N, A. Stankevich S, Menenti M. Evaluation of open Digital Elevation Models: estimation of topographic indices relevant to erosion risk in the Wadi M'Goun watershed, Morocco. *AIMS Geosci*. 2020;6(2):231-257. doi:10.3934/geosci.2020014

22. Tadono T, Nagai H, Ishida H, et al. Generation of the 30 M-MESH global digital surface model by alos prism. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. Vol 41. International Society for Photogrammetry and Remote Sensing; 2016:157-162. doi:10.5194/isprsarchives-XLI-B4-157-2016

23. Gajjar D, Darji K, Patel D, Prieto C, Han D. *A Comparative Study of Delineated Watersheds through ASTER, SRTM and ALOS for Evaluating Morphological Changes in Hathmati Basin, Gujarat, India*. Vol 20.; 2018.

24. AGREE - DEM Surface Reconditioning System. https://www.caee.utexas.edu/prof/maidment/gishydro/ferdi/research/agree/agree.html. Accessed January 27, 2021.

25. Li Z. Watershed modeling using arc hydro based on DEMs: a case study in Jackpine watershed. *Environ Syst Res*. 2014;3(1):11. doi:10.1186/2193-2697-3-11

26. ESRI. *Arc Hydro Tools - Tutorial*.; 2011. http://downloads.esri.com/archydro/ArcHydro/Tutorial/Doc/Arc Hydro Tools 2.0 - Tutorial.pdf.

27. About OpenStreetMap - OpenStreetMap Wiki. https://wiki.openstreetmap.org/wiki/About_OpenStreetMap. Accessed March 20, 2021.

28. Novel-T. Environmental Surveillance: Supporting Polio eradication. https://www.es.world/. Accessed April 20, 2021.

29. Environmental Sites Catalog. https://www.es.world/#!/catalog. Accessed February 2, 2021.

30. Price list | AW3D Global High-resolution 3D Map – Global Digital Elevation Model. https://www.aw3d.jp/en/pricelist/. Accessed October 18, 2020.

31. Bhutta ZA, Gaffey MF, Crump JA, et al. Typhoid Fever: Way Forward. *Am J Trop Med Hyg*. 99(3_Suppl):89-96. doi:10.4269/ajtmh.18-0111

32. World Health Organization. Typhoid. https://www.who.int/news-room/fact-sheets/detail/typhoid. Accessed February 9, 2021.

33. Crump JA, Luby SP, Mintz ED. The global burden of typhoid fever. *Bull World Health Organ*. 2004;82(5):346-353. doi:10.1590/S0042-96862004000500008

34. Sur D, von Seidlein L, Manna B, et al. The malaria and typhoid fever burden in the slums of Kolkata, India: data from a prospective community-based study. *Trans R Soc Trop Med Hyg*. 2006;100(8):725-733. doi:10.1016/j.trstmh.2005.10.019

35. Saha MR, Dutta P, Bhattacharya SK, et al. Occurrence of multi-drug resistant Salmonella typhi in Calcutta. *Indian J Med Res - Sect A Infect Dis*. 1992;95(JULY):179-180.

36. World Bank. *India - Vulnerability of Kolkata Metropolitan Area to Increased Precipitation in a Changing Climate*. Washington D.C.; 2011. http://documents.worldbank.org/curated/en/834561468041134833/India-Vulnerability-of-Kolkata-metropolitan-area-to-increased-precipitation-in-a-changing-climate.

37. Mukhopadhyay B, Sur D, Gupta S, Ganguly N. Typhoid fever: Control & challenges in India. *Indian J Med Res*. 2019;150(5):437-447. doi:10.4103/ijmr.IJMR_411_18

38. Moe CL. *The SaniPathApproach to Fecal Exposure Assessment and Application to Typhoid Transmission*.; 2016. https://www.coalitionagainsttyphoid.org/wp-content/uploads/2016/07/4-The-SaniPath-Approach-to-Fecal-Exposure-Assessment-and-Application-to-Typhoid-Transmission.pdf.

39. Danielson WC. The origin and development of the water carriage system of sewerage. 1910:29 leaves. file://catalog.hathitrust.org/Record/100180020.

40. Calcutta Metropolitan Planning Organization. Master Plan for Water Supply, Sewerage, and Drainage Calcutta Metropolitan District 1966-2001.

41. Basu N, Dey A, Ghosh D. Kolkata's brick sewer renewal: History, challenges and benefits. *ICE Proc Civ Eng*. 2013;166:74-81. doi:10.1680/cien.12.00016

42. Census of India. Government of India Website for Census.

43. Government of West Bengal. Irrigation & Waterways Department. https://wbiwd.gov.in/index.php/applications/kolkata_drainage. Accessed March 21, 2021.

44. Rainfall monitoring. http://www.meteocal.in/rain_flood.php. Accessed February 26, 2021.

45. Mukherjee J. *Blue Infrastructures : Natural History, Political Ecology and Urban Development in Kolkata*. Vol 1st ed. 20. Singapore: Springer; 2020. https://login.proxy.library.emory.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=2486075&site=ehost-live&scope=site.

46. COVID-19 Updates | Ghana. https://ghanahealthservice.org/covid19/. Accessed March 8, 2021.

47. Strengthening Contact Tracing for COVID-19 in Greater Accra Region | WHO | Regional Office for Africa. https://www.afro.who.int/news/strengthening-contact-tracing-covid-19-greater-accra-region. Accessed March 8, 2021.

48. Bonful HA, Addo-Lartey A, Aheto JMK, Ganle JK, Sarfo B, Aryeetey R. Limiting spread of COVID-19 in Ghana: Compliance audit of selected transportation stations in the Greater Accra region of Ghana. Di Gennaro F, ed. *PLoS One*. 2020;15(9):e0238971. doi:10.1371/journal.pone.0238971

49. World Bank Group. *Rising through Cities in Ghana Ghana Urbanization Review Overview Report*. World Bank, Washington, DC; 2014. www.worldbank.org. Accessed March 9, 2021.

50. World Bank Project : Greater Accra Resilient and Integrated Development Project - P164330. https://projects.worldbank.org/en/projects-operations/project-detail/P164330. Accessed April 2, 2021.

51. World Bank Group. *Enhancing Urban Resilience in the Greater Accra Metropolitan Area*. Washington D.C.; 2017. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/949241495793834492/enhancing-urban-resilience-in-the-greater-accra-metropolitan-area. Accessed March 9, 2021.

52. Gretsch SR, Ampofo JA, Baker KK, et al. Quantification of exposure to fecal contamination in open drains in four neighborhoods in Accra, Ghana. *J Water Health*. 2016;14(2):255-266. doi:10.2166/wh.2015.138

53. Robb K, Null C, Teunis P, Yakubu H, Armah G, Moe CL. Assessment of Fecal Exposure Pathways in Low-Income Urban Neighborhoods in Accra, Ghana: Rationale, Design, Methods, and Key Findings of the SaniPath Study. *Am J Trop Med Hyg*. 2017;97(4):1020-1032. doi:10.4269/ajtmh.16-0508

54. Emory University. Results Dashboard | SaniPath. https://www.sanipath.org/results-dashboard. Accessed April 18, 2021.

55. Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in the Netherlands. *Environ Sci Technol Lett*. 2020;7(7):511-516. doi:10.1021/acs.estlett.0c00357

56. WGS 84 / UTM zone 45N. https://georepository.com/crs_32645/WGS-84-UTM-zone-45N.html. Accessed March 13, 2021.

57. Kolkata_Bounds_Census_073119 (FeatureServer). https://services1.arcgis.com/4TXrdeWh0RyCqPgB/arcgis/rest/services/Kolkata_Bounds_Census_073119/FeatureServer. Accessed March 31, 2021.

58. Berendes DM, de Mondesert L, Kirby AE, et al. Variation in E. coli concentrations in open drains across neighborhoods in Accra, Ghana: The influence of onsite sanitation coverage and interconnectedness of urban environments. *Int J Hyg Environ Health*. 2020;224:113433. doi:10.1016/j.ijheh.2019.113433

59. WGS 84 / UTM zone 30N. https://georepository.com/crs_32630/WGS-84-UTM-zone-30N.html. Accessed March 17, 2021.

60. TREND Group. About Us. https://www.trendgroupgh.org/. Accessed April 19, 2021.

61. Moore S, Dongdem AZ, Opare D, et al. Dynamics of cholera epidemics from Benin to Mauritania.

PLoS Negl Trop Dis. 2018;12(4):e0006379. doi:10.1371/journal.pntd.0006379

62. GPS coordinates on Google Maps | Latitude Longitude Search. https://www.maps.ie/coordinates.html. Accessed April 1, 2021.

63. Sharma S, Jain A, Singh AK. *Watershed and Stream Network Delineation Using ArcHydro*.; 2015. https://www.researchgate.net/publication/286530037_Watershed_and_Stream_Network_Delineation_Using_ArcHydro. Accessed March 14, 2021.

64. Tarboton DG, Bras RL, Rodriguez-Iturbe I. On the extraction of channel networks from digital elevation data. *Hydrol Process*. 1991;5(1):81-100. doi:10.1002/hyp.3360050107

65. Kolkata Municipal Corporation. Official Website of Kolkata Municipal Corporation. https://www.kmcgov.in/KMCPortal/jsp/PumpingStation.jsp. Accessed March 23, 2021.

TABLES

Table 1. GPS location, altitude, and installed capacity of sampled pumping stations sampled for ES of *S*. Typhi and *S*. Paratyphi A in Kolkata, India

Pumping Station	Latitude	Longitude	Altitude (m)	Installed Capacity* (cusecs)
Ambedkar	22.5356091°N	88.3992867°E	14	NA
Baishnabghat	22.4710588°N	88.3924983°E	7	NA
Ballygunge	22.5361962°N	88.3746299°E	3	1,233
Bangur	22.5066961°N	88.3567244°E	12	NA
Chingrighata	22.5574774°N	88.3920952°E	7	100
Cossipore	22.6060572°N	88.3734558°E	7	NA
Dhapa Lock	22.5576242°N	88.4110281°E	7	480
Duttabagan	22.6078184°N	88.3923887°E	5	48
Jorabridge	22.4956886°N	88.3913613°E	7	NA
Kulia Tangra	22.5580645°N	88.3894534°E	8	40
Pagladanga	22.5561565°N	88.3959111°E	7	48
Palmer's Bridge	22.5632013°N	88.3777120°E	7	1,184
Topsia	22.5408928°N	88.3868116°E	7	65

NA = Not Available
*Installed capacities from Government of West Bengal, Irrigation & Waterways Department[43]

Table 2. Values input in fields in the DEM Reconditioning tool in Arc Hydro Pro Tools compared to the default settings.

	Number of cells for stream buffer	Smooth drop in Z units	Sharp drop in Z units
Small waterways	1	1	10
Medium waterways	2	5	10
Large waterways	5	3	10
Default Settings	5	10	1,000

Table 3. Number of samples collected and estimated geographic catchment and population size for sampled pumping stations in Kolkata.

Pumping Station	No. Samples Collected	Geographic Catchment Area (sq km)	Population Size Estimate
Kulia Tangra	20	0.12	1,493
Bangur	5	0.06	1,525
Jorabridge	5	0.13	3,323
Baishnabghat	5	0.59	5,916
Chingrighata	22	0.78	9,405
Pagladanga	30	1.13	13,160
Topsia	31	1.19	33,542
Ambedkar	31	1.19	34,517
Duttabagan	5	1.37	49,788
Cossipore	4	3.10	95,471
Ballygunge	20	5.71	225,074
Dhapa Lock	19	12.19	378,753
Palmer's Bridge	35	20.42	890,681
Total	**232**	**47.98**	**1,742,649**

Table 4. Summary statistics for estimated geographic area and population size of catchment areas upstream of sampled pumping station for Kolkata.

	Minimum	Maximum	Average	Std. Dev.
Geographic Area (sq. km.)	0.06	20.42	3.69	6.04
Population Size	1,493	890,681	134,050	252,990

Table 5. Wards identified within upstream catchment areas of sampled pumping stations and the percentage of the total ward area that is contained within the catchment area.

Pumping Station	Wards Identified Within Catchment Area	Percentage of Total Ward Area Contained within PS Catchment Area
Kulia Tangra	57	3
	58	0.2
Bangur	93	3
Jorabridge	104	8
	106	2
Baishnabghat	109	0.4
	110	21
Chingrighata	57	6
	58	6
	66	1
Pagladanga	57	17
	58	5
Topsia	55	1
	56	18
	58	5
	59	21
	66	6
Ambedkar	58	0.1
	59	0.4
	66	35
Duttabagan	2	7
	3	45
	4	63
	5	2
Cossipore	1	100
	6	100
Ballygunge*	54 - 55, 60, 64, 68	100
	65	69
	69	33
	85	1
	86	20
Dhapa Lock*	13 - 14, 29 - 35	100
	5	55
Palmer's Bridge*	7 - 12, 15 - 28, 36 - 53, 61 - 63	100

*All or some of wards assumed to be contributing 100% of ward's area based on available sewer diagrams[44] and KMC source (2005) cited in Mukherjee (2020)[45]

Table 6. Distribution of wards by pumping station catchment area.

Pumping Station	Wards Identified Within PS Catchment Area	Percentage of PS Catchment Area by Ward
Kulia Tangra	57	84
	58	16
Bangur	93	100
Jorabridge	104	77
	106	23
Baishnabghat	109	7
	110	93
Chingrighata	57	31
	58	64
	66	5
Pagladanga	57	62
	58	38
Topsia	55	1
	56	10
	58	38
	59	34
	66	17
Ambedkar	58	0.5
	59	1
	66	99
Duttabagan	2	9
	3	45
	4	44
	5	2
Cossipore	1	44
	6	56
Ballygunge	54	7
	55	18
	60	10
	64	15
	65	17
	68	16
	69	12
	85	0.2
	86	3
Dhapa Lock	5	8
	13	8
	14	10
	29	5
	30	6
	31	14
	32	16
	33	18

	34	8
	35	9
Palmer's Bridge	7	2
	8	2
	9	1
	10	2
	11	2
	12	3
	15	3
	16	2
	17	2
	18	1
	19	1
	20	1
	21	2
	22	2
	23	1
	24	1
	25	2
	26	2
	27	2
	28	2
	36	5
	37	2
	38	2
	39	1
	40	2
	41	1
	42	2
	43	1
	44	3
	45	8
	46	6
	47	2
	48	1
	49	1
	50	2
	51	1
	52	1
	53	2
	61	3
	62	2
	63	19

Table 7. Proposed strategic sample locations and the estimated geographic area and population size captured by these locations in Accra.

Strategic Sampling Point	Latitude	Longitude	Geographic Area Served (sq km)	Population Size Estimate
1	5.5371562°N	0.2559009°W	3.62	61,885
2	5.5948968°N	0.2421168°W	3.04	78,410
3	5.5867519°N	0.2164054°W	23.25	153,280
4	5.5738117°N	0.2044346°W	6.01	152,001
5	5.5488370°N	0.2255134°W	14.30	253,611
6	5.5435215°N	0.2218098°W	3.21	48,489
7	5.5649079°N	0.1776306°W	1.63	17,032
Total	-	-	55.05	764,708

Table 8. Summary statistics for estimated geographic area and population size of catchment areas upstream of optimal sampling locations for Accra.

	Minimum	Maximum	Average	Std. Dev.
Geographic Area (sq km)	1.63	23.25	7.86	7.99
Population Size	17,032	253,611	109,244	81,647

Table 9. Accra neighborhoods identified within upstream catchment areas of strategic sampling locations and the percentage of the total neighborhood area that is contained within the catchment area.

Strategic Sampling Point	Neighborhoods Within Catchment Area	Percentage of Total Neighborhood Area Contained within Catchment Area
1	Dansoman	29
	Mamponse	15
	New Mamprobi	10
	Russia	40
	South Odorkor	0.2
	Sukura	42
	West Abbossey Okai	15
2	Abeka	54
	Abeka Lapaz	5
	Bubuashie	5
	Darkuman	36
	New Fadama	23
	Nii Boi Town	6
3	Abelemkpe	51
	Abofu	1
	Achimota	45
	Airport Residential Area	64
	Alajo	62
	Anumle	18
	Dzorwulu	100
	East Legon	31
	Kotobabi	100
	Kpehe	84
	Legon	35
	Legon Staff Village	49
	Mamobi	11
	Mempeasem	8
	New Town	40
	Okponglo	100
	Roman Ridge	93
4	Airport Residential Area	36
	Kanda	47
	Kokomlemle	22
	Kotobabi	0.02
	Mamobi	89
	New Town	43
	Nima	73
	Roman Ridge	7

5	Abbossey Okai	100
	Bubuashie	85
	Dansoman	0.01
	Darkuman	56
	Kaneshie	84
	Korle-Bu	9
	Kwashieman	12
	Lartebiokorshie	23
	Mataheko	100
	North Kaneshie	15
	North Odorkor	15
	Russia	20
	Sabon Zongo	100
	South Odorkor	24
	West Abbossey Okai	85
	Zoti	81
6	Accra Central	63
	Adabraka	6
	Adedenkpo	38
	Korle Dudor	80
	Ridge	38
	Tudu	50
	Ussher Town	1
7	Kanda	2
	Osu	24
	Ridge	3
	Ringway	89

Table 10. Distribution of neighborhoods for each strategic sampling point's catchment area.

Strategic Sampling Point	Neighborhoods Identified Within Catchment Area	Percentage of Strategic Sampling Point Catchment Area by Neighborhood
1	Dansoman	51
	Mamponse	7
	New Mamprobi	2
	Russia	13
	South Odorkor	0.2
	Sukura	16
	West Abbossey Okai	10
2	Abeka	45
	Abeka Lapaz	1
	Bubuashie	5
	Darkuman	43
	New Fadama	6
	Nii Boi Town	0.5
3	Abelemkpe	4
	Abofu	0.1
	Achimota	13
	Airport Residential Area	11
	Alajo	5
	Anumle	2
	Dzorwulu	16
	East Legon	8
	Kotobabi	6
	Kpehe	1
	Legon	17
	Legon Staff Village	4
	Mamobi	1
	Mempeasem	1
	New Town	3
	Okponglo	1
	Roman Ridge	7
4	Airport Residential Area	24
	Kanda	17
	Kokomlemle	6
	Mamobi	21
	New Town	11
	Nima	18
	Roman Ridge	2
5	Abbossey Okai	5
	Bubuashie	16
	Darkuman	14

	Kaneshie	16
	Korle-Bu	1
	Kwashieman	2
	Lartebiokorshie	3
	Mataheko	6
	North Kaneshie	2
	North Odorkor	2
	Russia	2
	Sabon Zongo	3
	South Odorkor	7
	West Abbossey Okai	14
	Zoti	7
6	Accra Central	18
	Adabraka	4
	Adedenkpo	13
	Korle Dudor	33
	Ridge	25
	Tudu	7
	Ussher Town	0.1
7	Kanda	3
	Osu	40
	Ridge	4
	Ringway	53

Table 11. Comparison of estimated catchment sizes (geographic area and population) of sampled pumping stations for ES in Kolkata between Novel-T and the mixed-methods model developed through this thesis project.

	Geographic area (sq km)		Population Size	
Pumping Station	**Novel-T**	**Mixed-Methods Model**	**Novel-T**	**Mixed-Methods Model**
Ambedkar	<1	1.19	2,262	34,517
Baishnabghat	NA	0.59	NA	5,916
Ballygunge	<1	5.71	8,199	225,074
Bangur	NA	0.06	NA	1,525
Chingrighata	18.87	0.78	892,646	9,405
Cossipore	NA	3.10	NA	95,471
Dhapa Lock	2.07	12.19	59,890	378,753
Duttabagan	NA	1.37	NA	49,788
Jorabridge	NA	0.13	NA	3,323
Kulia Tangra	<1	0.12	57	1,493
Pagladanga	<1	1.13	7,395	13,160
Palmer's Bridge	<1	20.42	16,087	890,681
Topsia	<1	1.19	3,232	33,542

NA = Not Available

FIGURES

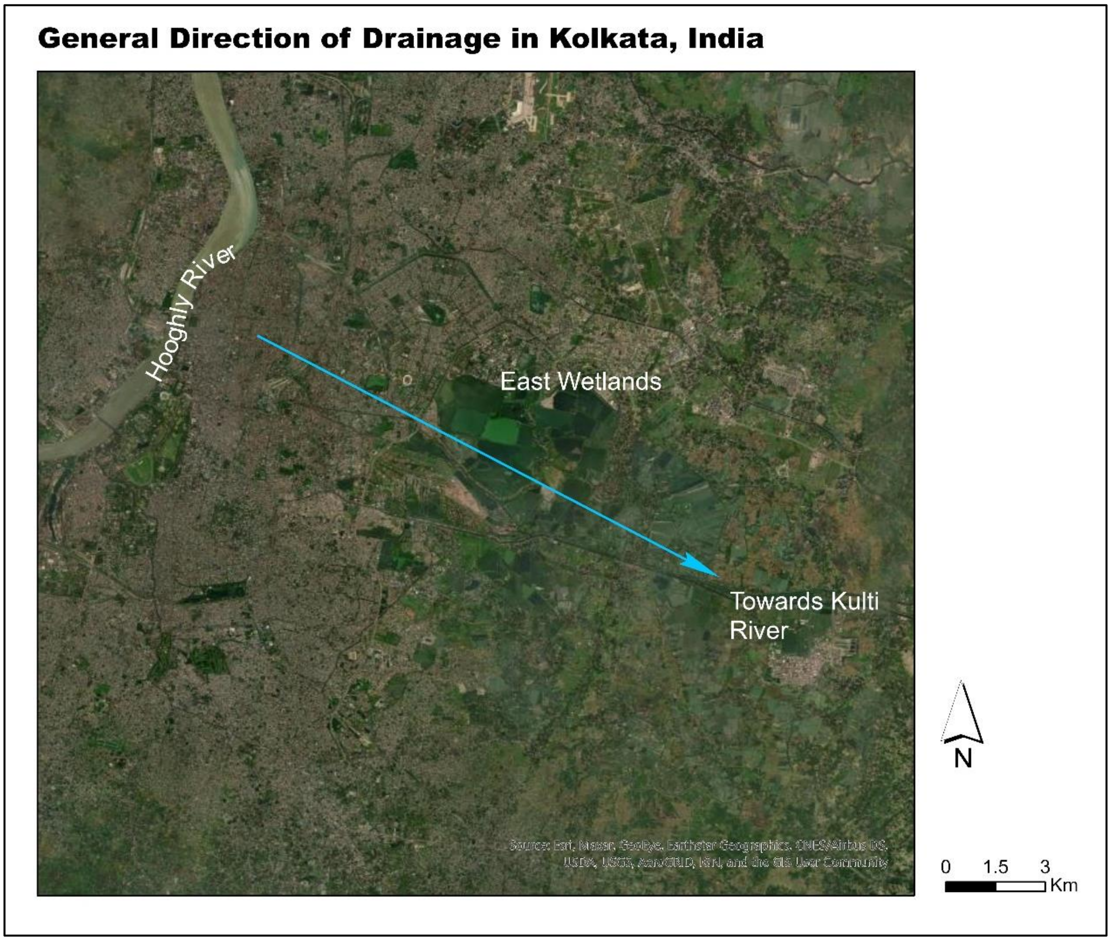


Figure 1. General direction of drainage is in southeasterly direction, from the Hooghly River through the East Kolkata Wetlands towards the Kulti River and, eventually, into the Bay of Bengal.

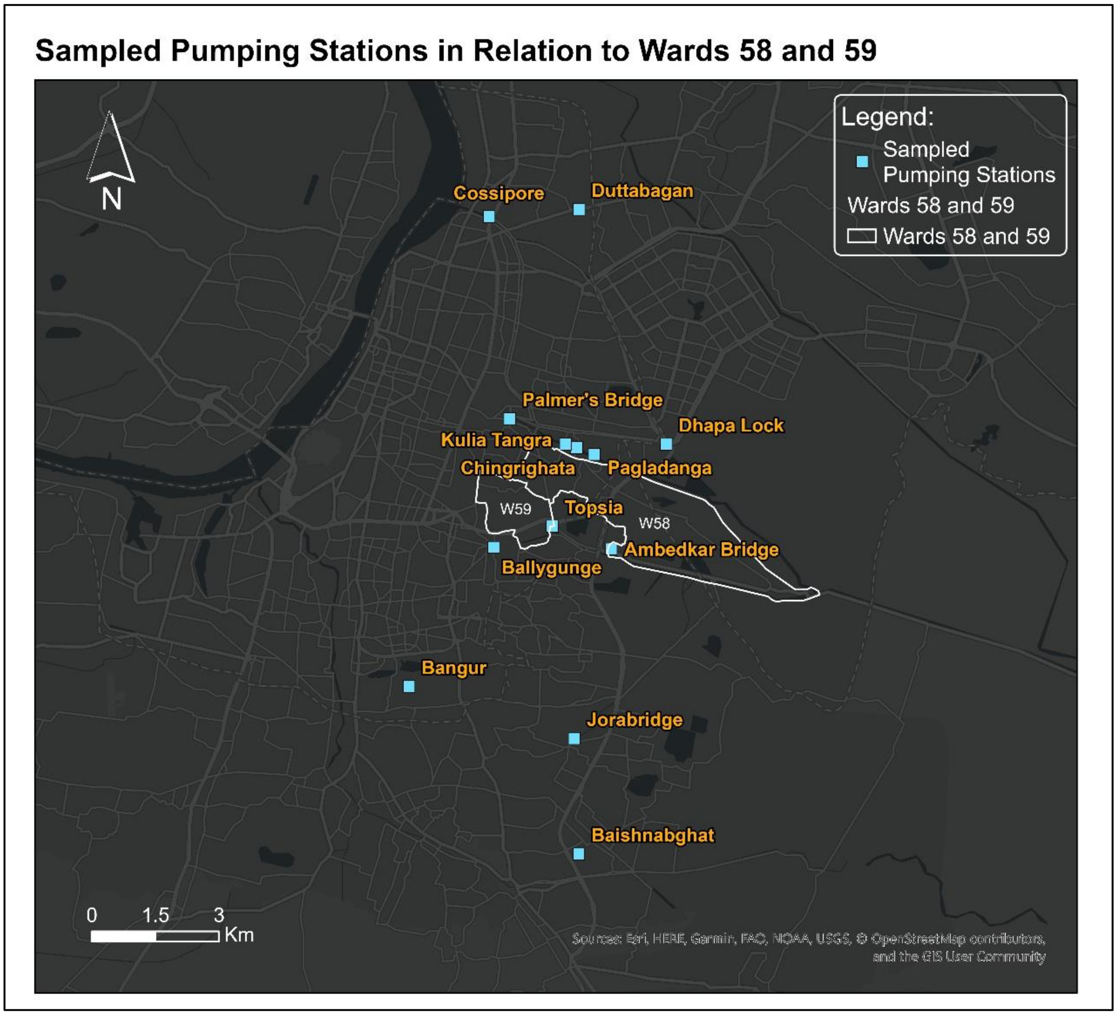


Figure 2. Locations of pumping stations sampled for ES for *S*. Typhi and *S*. Paratyphi A in relation to Wards 58 and 59, where active surveillance of clinical symptoms of typhoid fever was conducted. Selected pumping stations for ES were concentrated in/near these wards to supplement active surveillance.

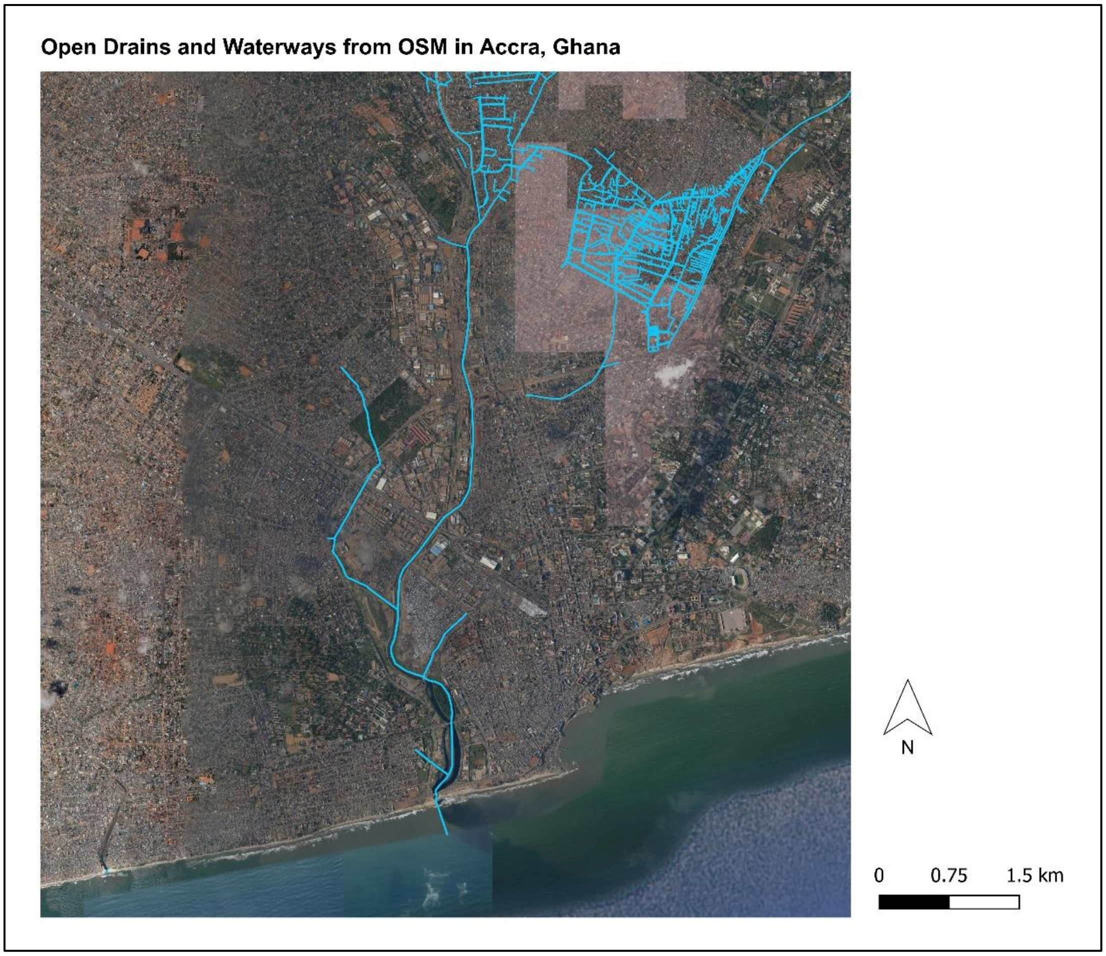


Figure 3. Open Drains and Waterways from OSM for Accra, Ghana. The open drains shown in the northeast corner of the map were not available for all of Accra from OpenStreetMap (OSM). However, providing a snapshot of these drains shows how interwoven they are in the urban landscape and how they drain into larger waterways that ultimately lead to the ocean.

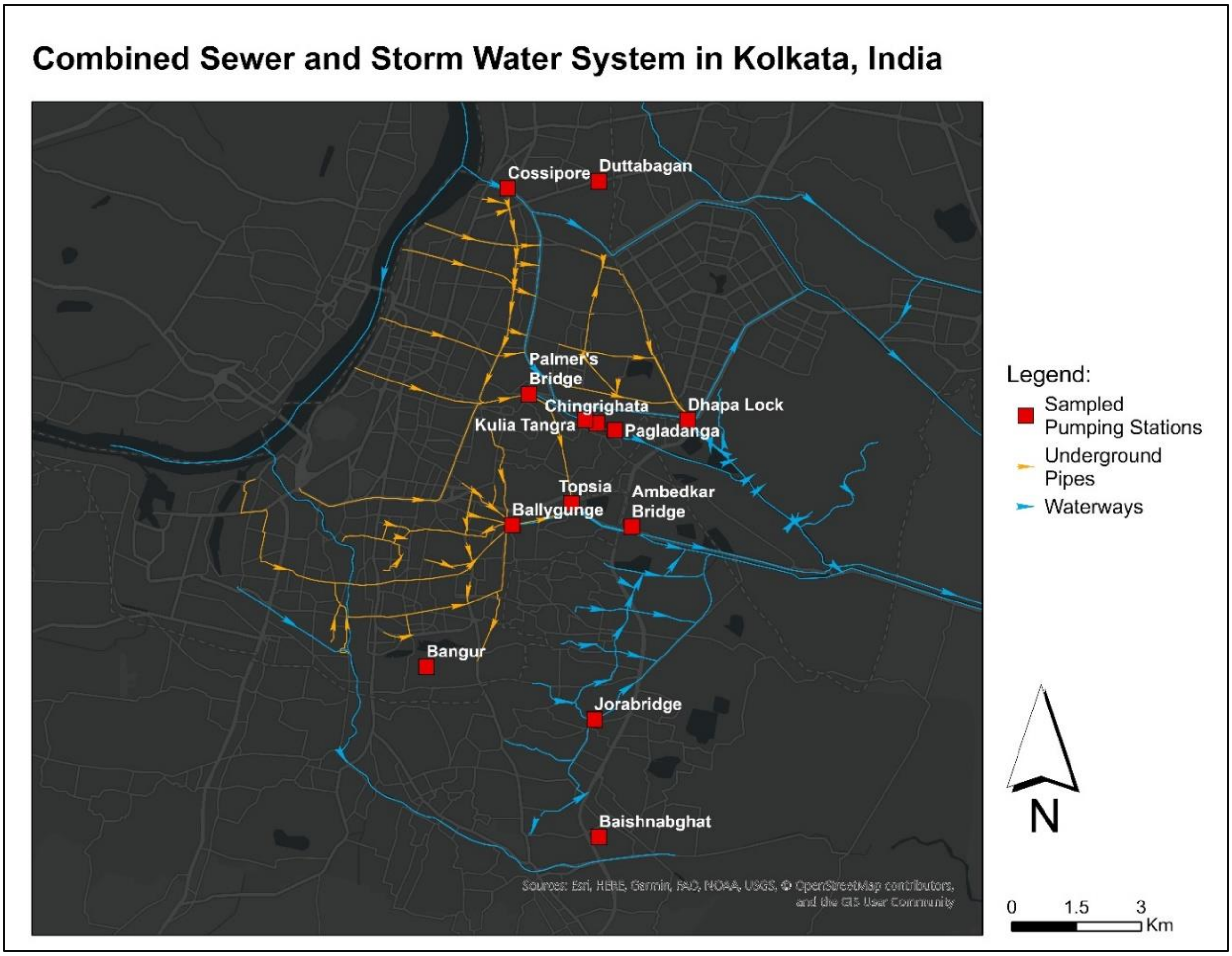


Figure 4. Connections between underground sewer and storm water pipes, open waterways, and sampled pumping stations were pieced together to better understand the movement of sewage and storm water throughout Kolkata, India.

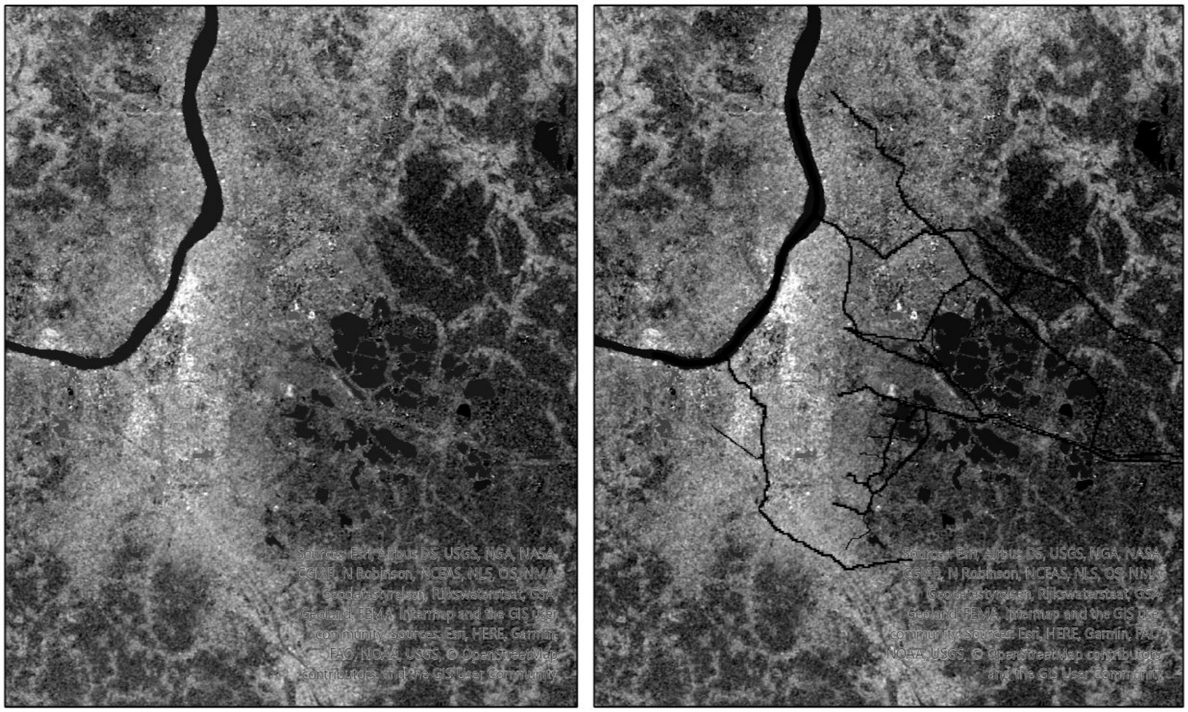


Figure 5. Before DEM Reconditioning tool was applied (left) and after DEM Reconditioning tool was applied. The black lines burned into the DEM in the reconditioned map represent artificial waterways (i.e., man-made canals) that were not captured through satellite imagery.

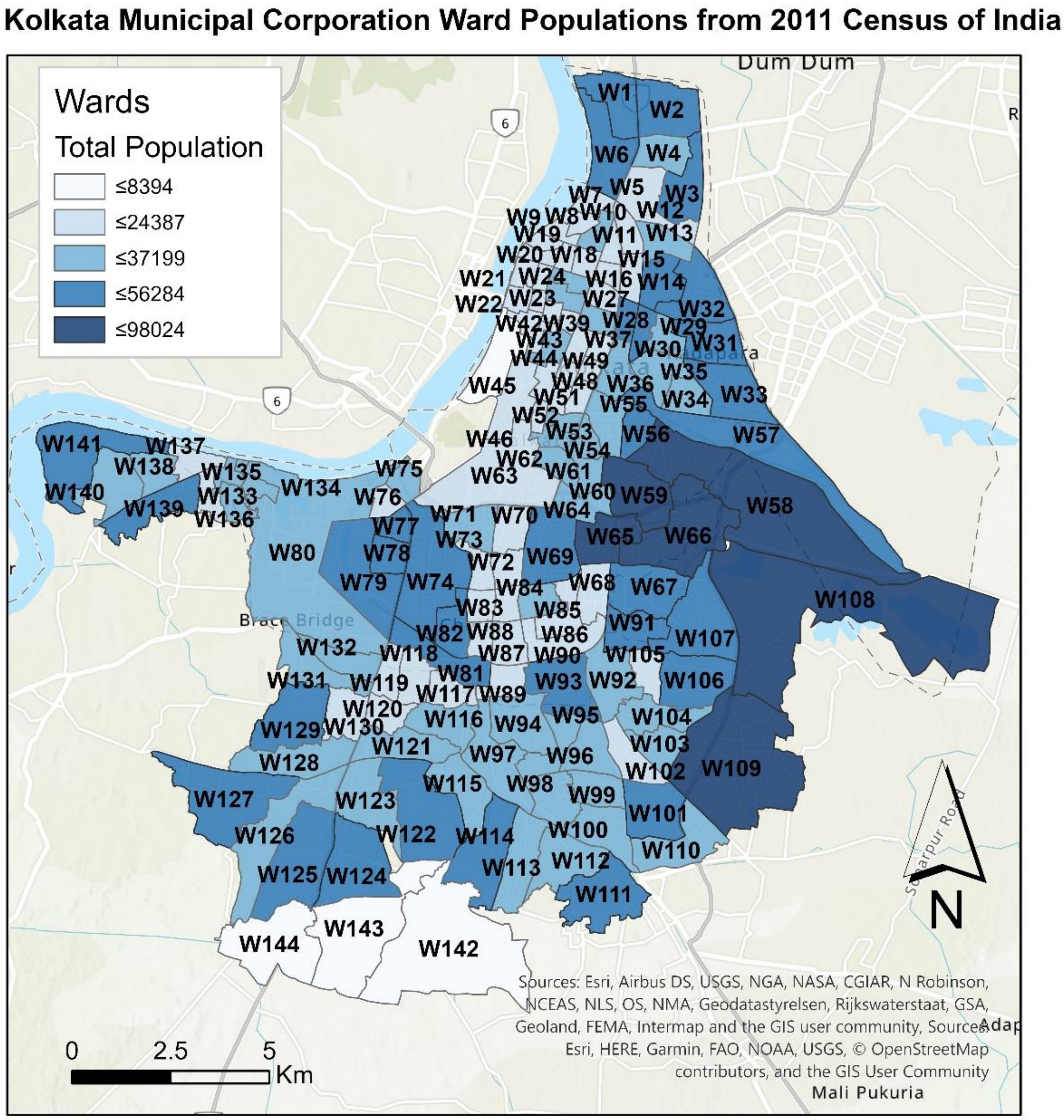
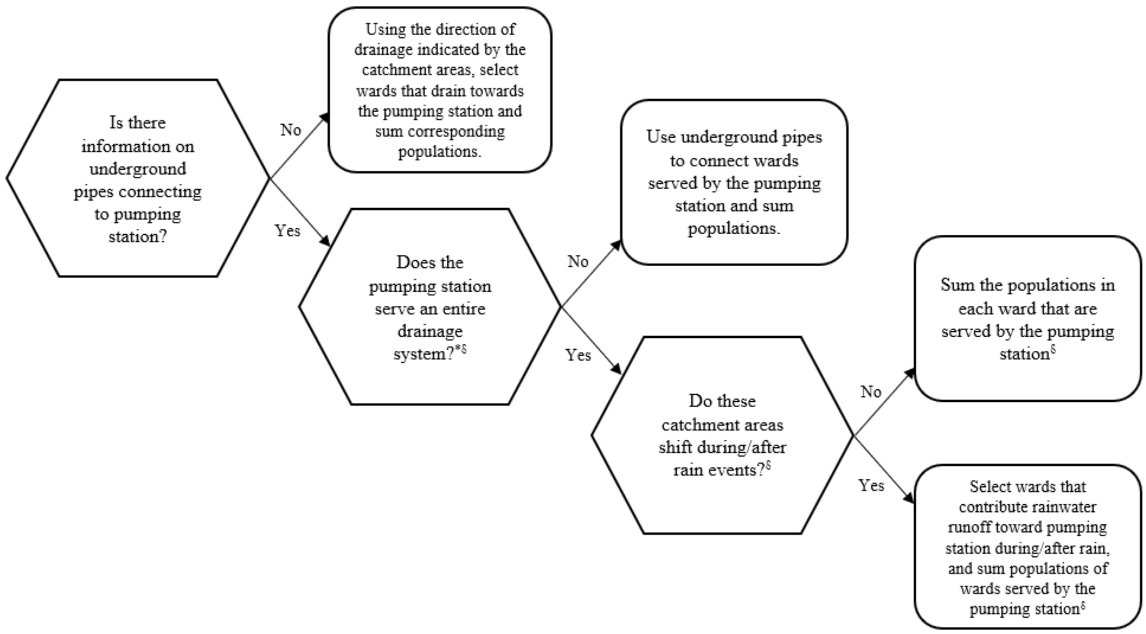


Figure 6. Total population per ward from 2011 Census of India[42] using Natural Breaks (Jenks) classification



* Town, Suburban, and Maniktala systems are served by Palmer's Bridge, Ballygunge, and Dhapa Lock pumping stations (respectively) [41]
§ Information obtained from *Blue Infrastructures: Natural History, Political Ecology and Urban Development in Kolkata* [41]

Figure 7. Process Flow Diagram for Estimating Catchment for Sampled Pumping Stations in Kolkata, India

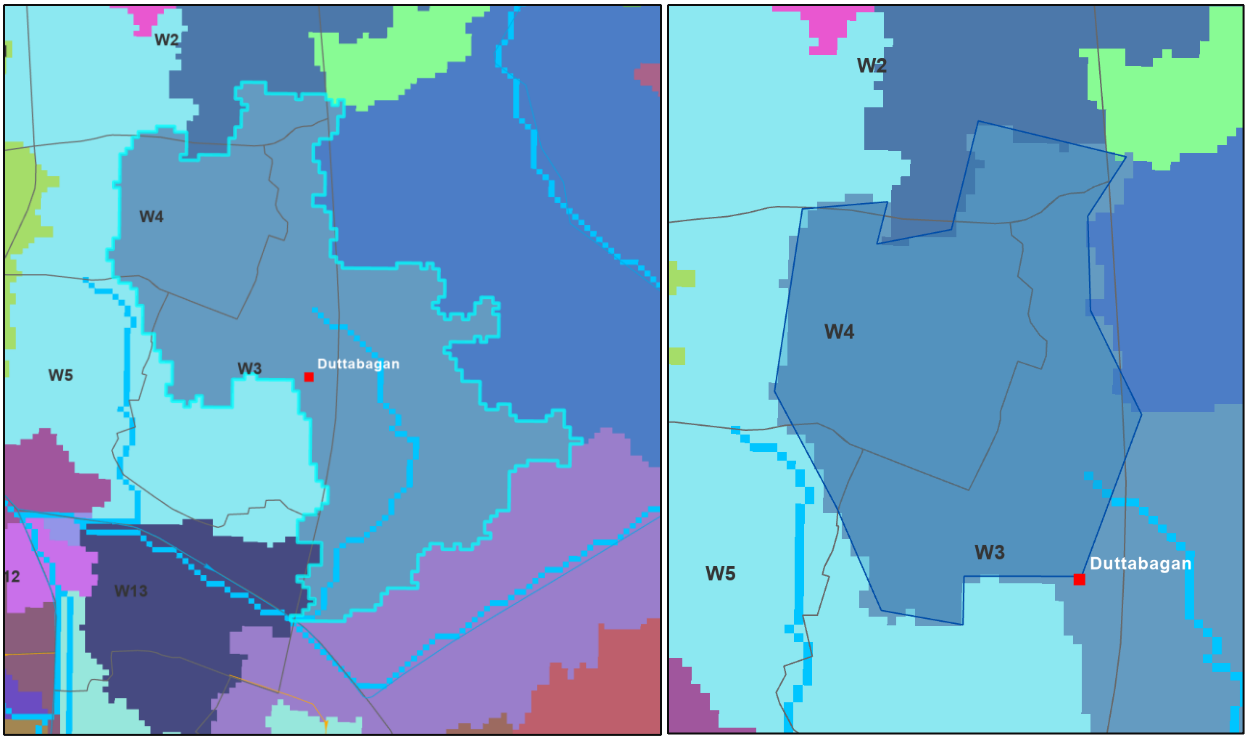


Figure 8a (left). Watershed in which Duttabagan PS is situated. **Figure 8b** (right). Perceived zone of catchment area that drains to the Duttabagan PS.

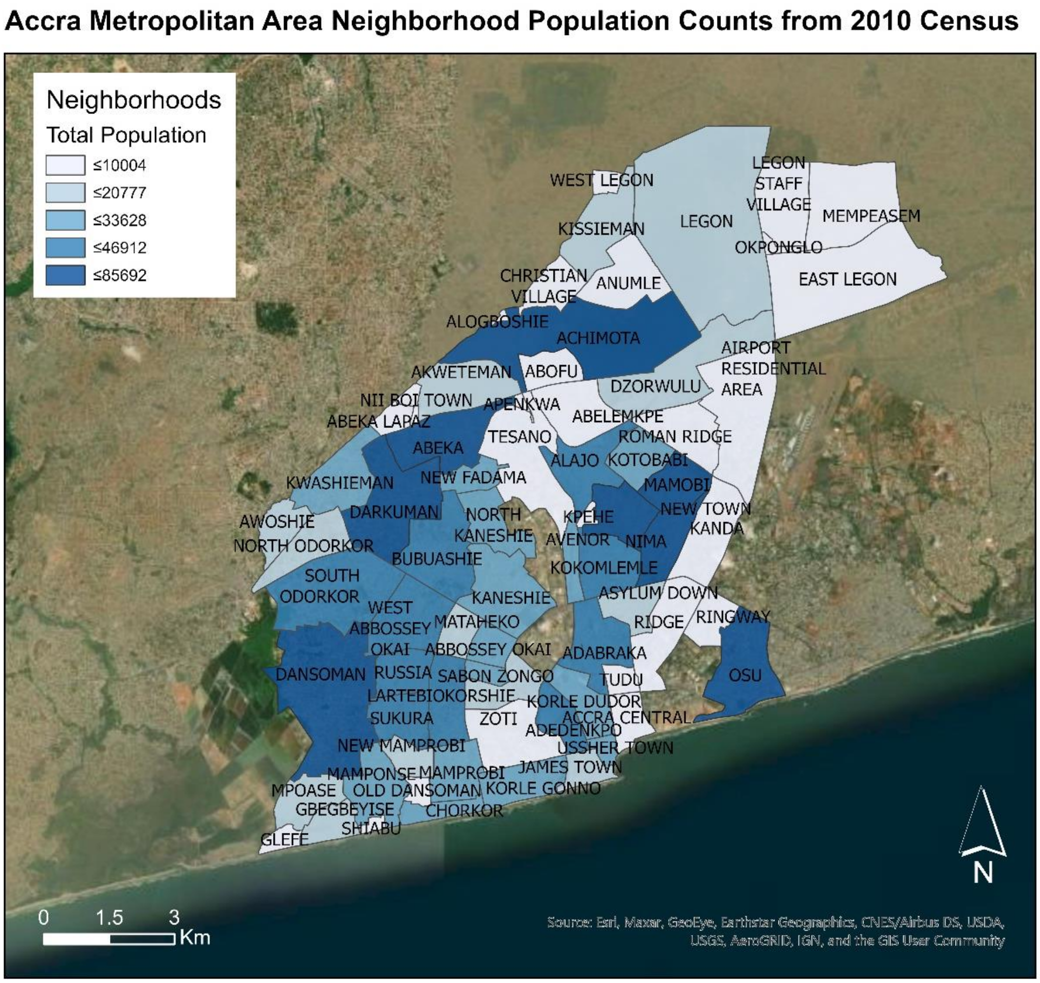


Figure 9. Total population per neighborhood from 2010 Census using Natural Breaks (Jenks) classification (shapefile used to create map provided by TREND Group in Ghana).

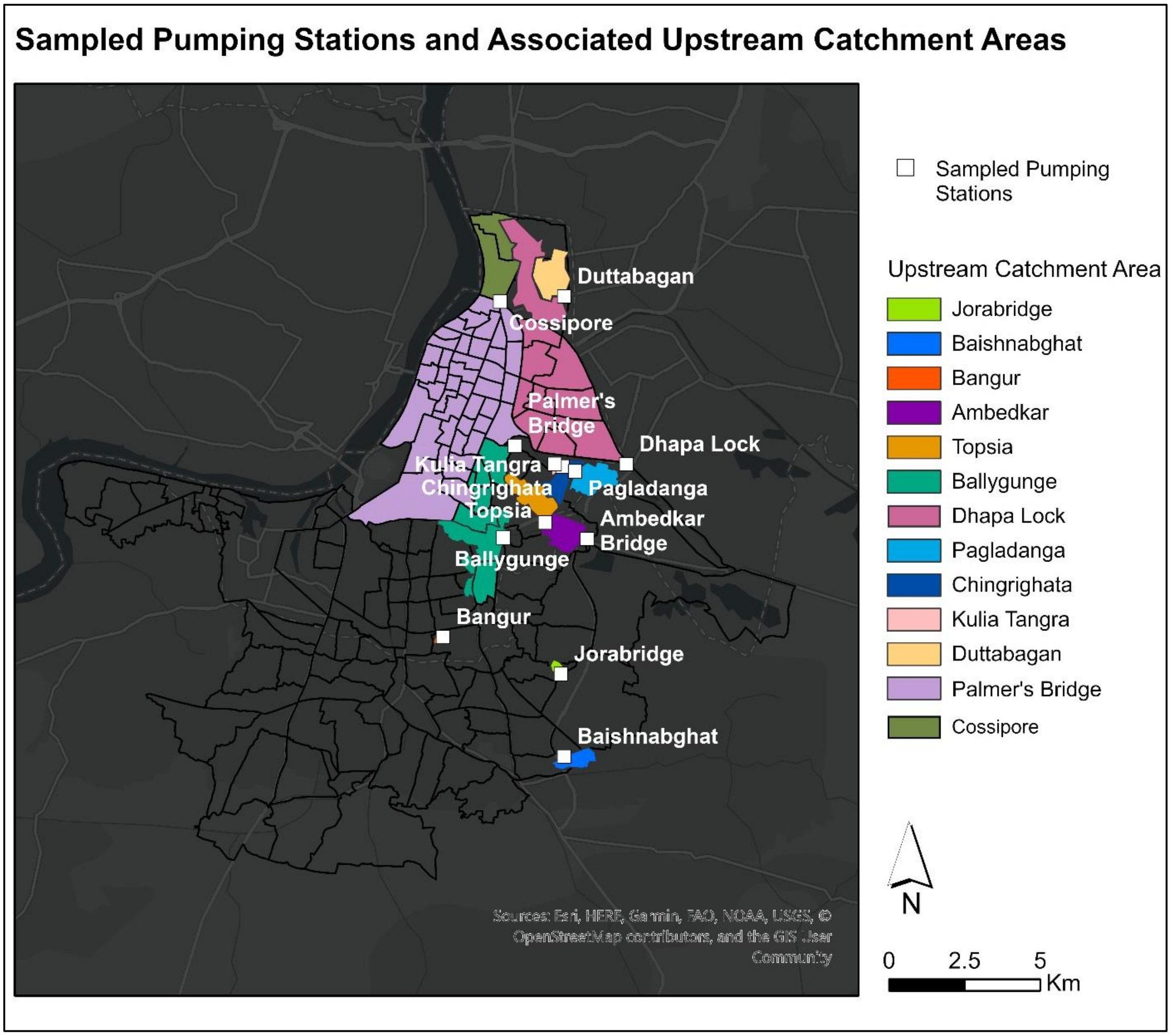


Figure 10. Mixed-methods, hybrid model used to map the upstream catchment areas of sampled pumping stations in Kolkata, India.

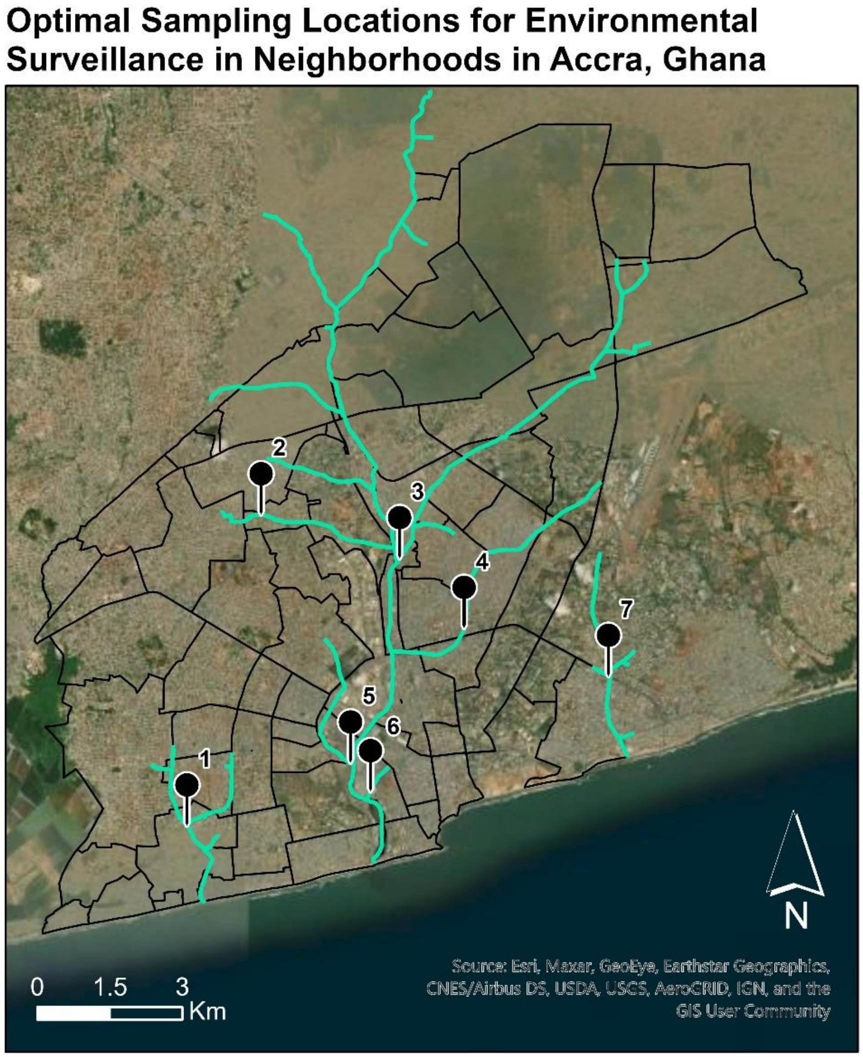


Figure 11. Seven strategic locations to collect samples for ES for SARS-CoV-2 in Accra, Ghana were identified and are shown here.

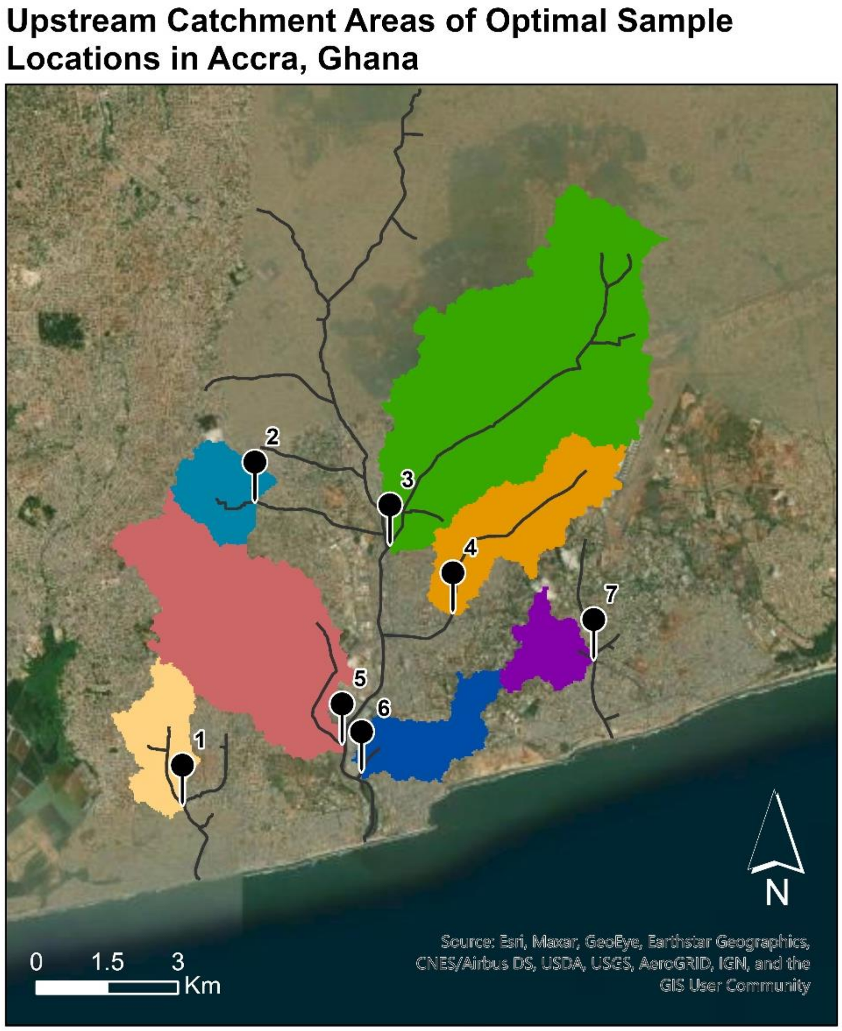


Figure 12. Upstream catchment areas of optimal locations to collect samples for ES for SARS-CoV-2 in Accra, Ghana.

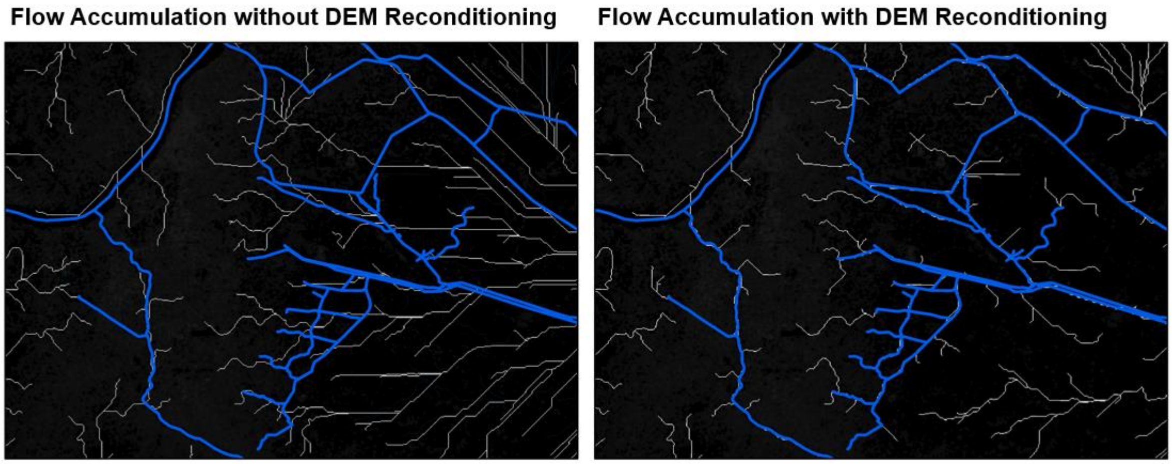


Figure 13. Comparing the drainage network calculated by the Flow Accumulation tool on a DEM without (left) and with (right) DEM Reconditioning. The outputs demonstrate that the DEM Reconditioning tool results in a drainage network that is a more accurate reflection of the true network of waterways.

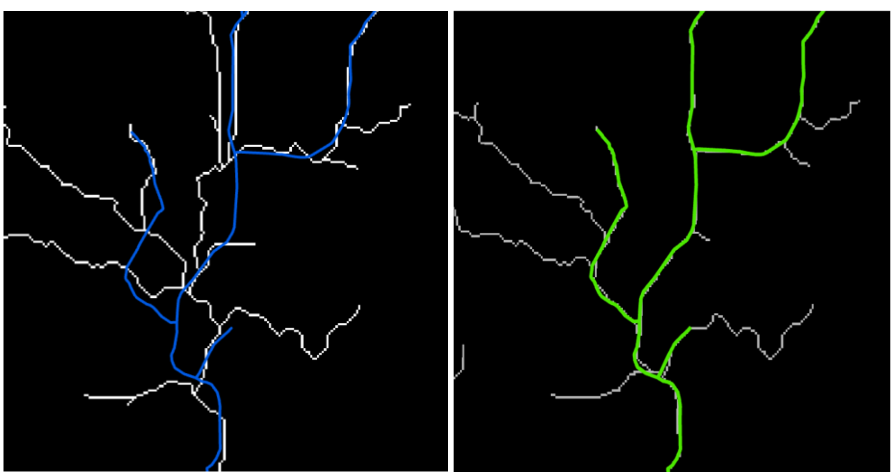


Figure 14. Before and after comparison for DEM Reconditioning step.

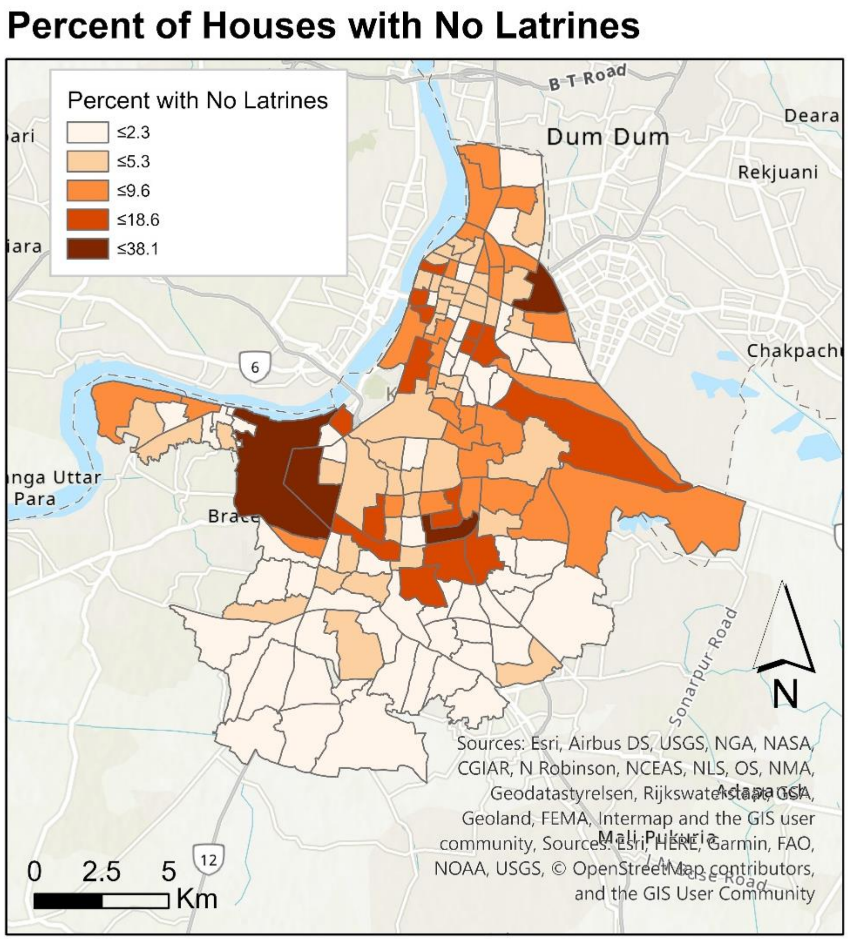


Figure 15. Percent of houses with no latrines in KMC wards according to 2011 Census of India.

APPENDICES

APPENDIX A: Step-by-Step Methodology for Creating Mixed-Methods Catchment Area Model for Kolkata, India

Methodology acknowledgement

The following resources and tutorials were referenced to guide development of the methodology:

- “Arc Hydro Geoprocessing Tools – Tutorial Version 2.0” by ESRI[26]
- “Watershed and Stream Network Delineation Using Arc Hydro” by Sudhakar Sharma, Akshay Jain, and Anupam Kumar Singh[63]
- “Exercise 4: Watershed and Stream Network Delineation” prepared by David G. Tarboton and David R. Maidment from the University of Texas[64]
- “Watershed Modeling Using Arc Hydro Based on DEMs: A Case Study in Jackpine Watershed” by Zhong Li[25]

Pre-requisite Steps

1. Install ArcGIS Pro
   1.1. Oftentimes, a university or organization has purchased an ArcGIS license, and this subscription can be extended to a user at the institution. To perform this thesis project, Emory University’s ArcGIS Pro license was extended and installed onto a personal computer. ArcMap can also be used to perform watershed delineation and the other mapping tasks discussed in this methodology. For those who do not have access to an ArcGIS license through their employer or educational institution, however, installation is rather costly. It is possible to perform this methodology using QGIS, a free, open-access GIS application, and this software should be explored to make this methodology more accessible to users around the world.
   1.2. To install ArcGIS Pro, visit the ESRI website (https://www.esri.com/en-us/home) to learn about various packages and pricing details. From this website, ArcGIS Pro can be downloaded and installed for use depending on the person’s or institution’s preference.
2. Install Arc Hydro extension

2.1. ESRI provides several versions of Arc Hydro to match different ArcGIS versions. To download the Arc Hydro extension, visit http://downloads.esri.com/archydro/archydro/. For this thesis, the ArcGIS Pro extension was downloaded and installed.

Building the Model in GIS

1. Preparing Waterways Shapefile

1.1. To extract waterways from OpenStreetMap (OSM), a query was developed and run in Overpass Turbo, which is a data mining tool for OSM[27]. This was performed in QGIS using the QuickOSM plugin. From the plugin, the "Quick query" option was selected, and the fields were filled out as shown in Figure A.1. The "Layer Extent" option was then selected, which executes the query to the extent of the geographic area framed in the map. Under the "OSM File" tab, the points option was deselected (points are not necessary for this task). Next, "Show query" was selected, which reveals the code shown in Figure A.2. The query was executed by selecting "Run query". The output from this operation is shown in Figure A.3.

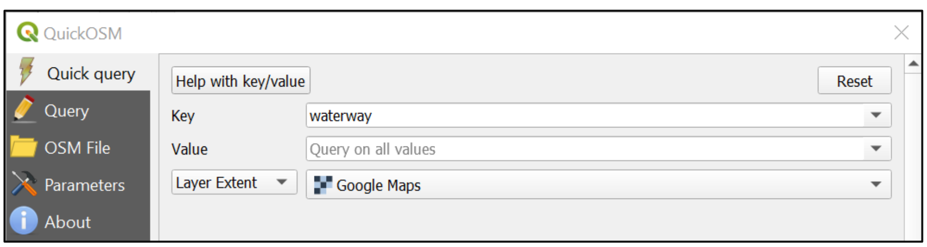


Figure A.1. Inputs into "Quick query" within the QuickOSM plugin in QGIS. This query extracts features tagged as "waterway" from the OSM database to reduce time spent manually creating these layers.

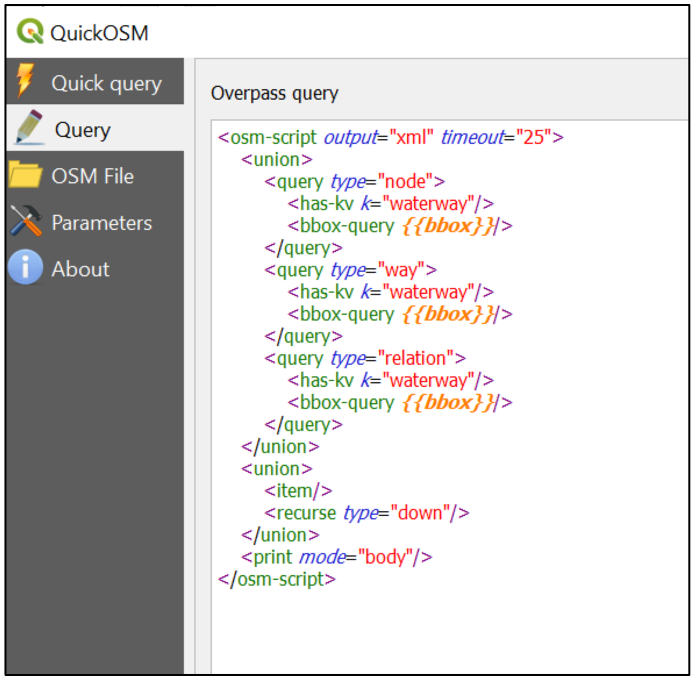


Figure A.2. The query that is run through the Overpass Turbo data mining tool for OSM.

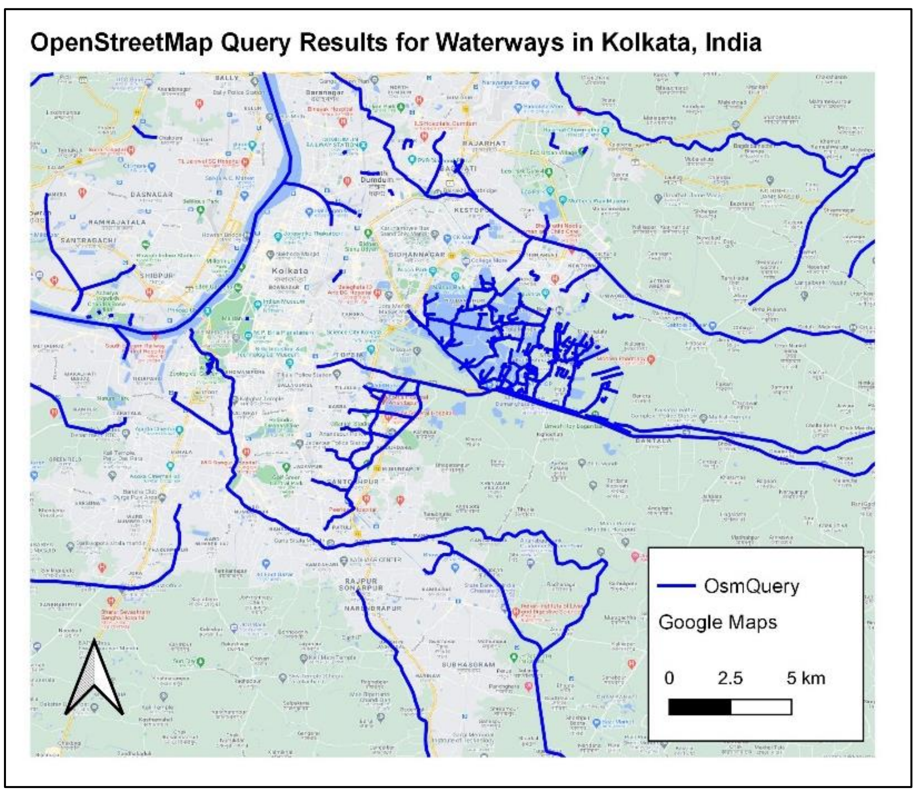


Figure A.3. Waterways extracted from OSM for the Kolkata area.

1.2. Several waterways were missing or incomplete. This was addressed by manually tracing missing waterways or extending incomplete lines over a Google Maps basemap. This process can be relatively time-consuming (depending on the geographic extent and quality of the query results); however, it is made significantly easier by the outputs of the OSM query.

1.3. Waterways were then categorized into three groups based on their widths, which were approximated using the measurement tool in Google Maps: small (width </= 12m); medium (12m < width </= 50m); large (> 50m). The waterways shapefile was then broken out into three shapefiles to create a feature for small, medium, and large streams.

1.4. These features were converted into raster format using the tool "Feature to Raster" and projected to the Coordinate Reference System (CRS) "WGS 1984 / UTM Zone 45N". This CRS was selected due to Kolkata's location between 84°E and 90°E in the northern hemisphere above 84°N [56].

2. Downloading DEMs from USGS EarthExplorer

2.1. SRTM 1-Arc Second Global DEM data was downloaded from the website https://earthexplorer.usgs.gov/. It is free to create an account, but this step must be performed before data can be downloaded.

2.2. A polygon was drawn around the area of interest that encompassed Kolkata. Next, the "Data Sets" tab was selected, then the following selections were made from the drop-down list: Digital Elevation → SRTM → SRTM 1 Arc-Second Global (Figure A.4). After selecting this option, the "Results" tab was clicked, and the DEM was then downloaded as a GeoTIFF file.

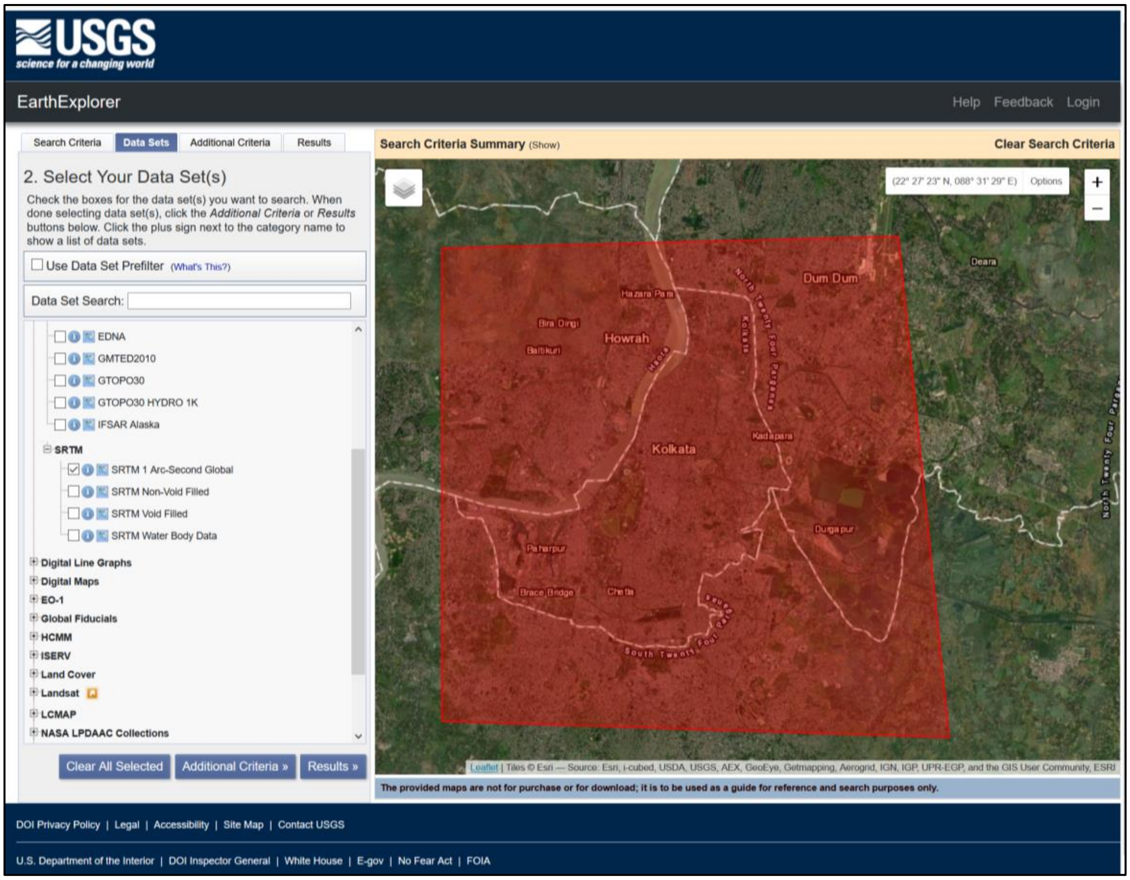


Figure A.4. Process for downloading SRTM 30-m DEM from USGS EarthExplorer

3. Add DEM in ArcGIS Pro

    3.1. The DEM was projected to CRS "WGS 1984 / UTM Zone 45N" using the "Project Raster" tool to unify layers under the same CRS.

4. DEM Reconditioning

    4.1. The DEM Reconditioning tool was selected from within Arc Hydro toolbox. This step was performed first on the small waterways raster and then performed twice more to burn the medium and large rasters into the DEM. These steps were performed iteratively (i.e., the output from the previous execution is used as the input in the next execution) to burn in waterways according to their size category to create a more realistic reconditioned DEM. The results of the process are shown in Figure A.5. The following values were entered for the fields “Number of cells for stream buffer”, “Smooth drop in Z units”, and “Sharp drop in Z units” (respectively) to achieve this:

- Small waterways: 1, 1, 10
- Medium waterways: 2, 5, 10
- Large waterways: 5, 3, 10

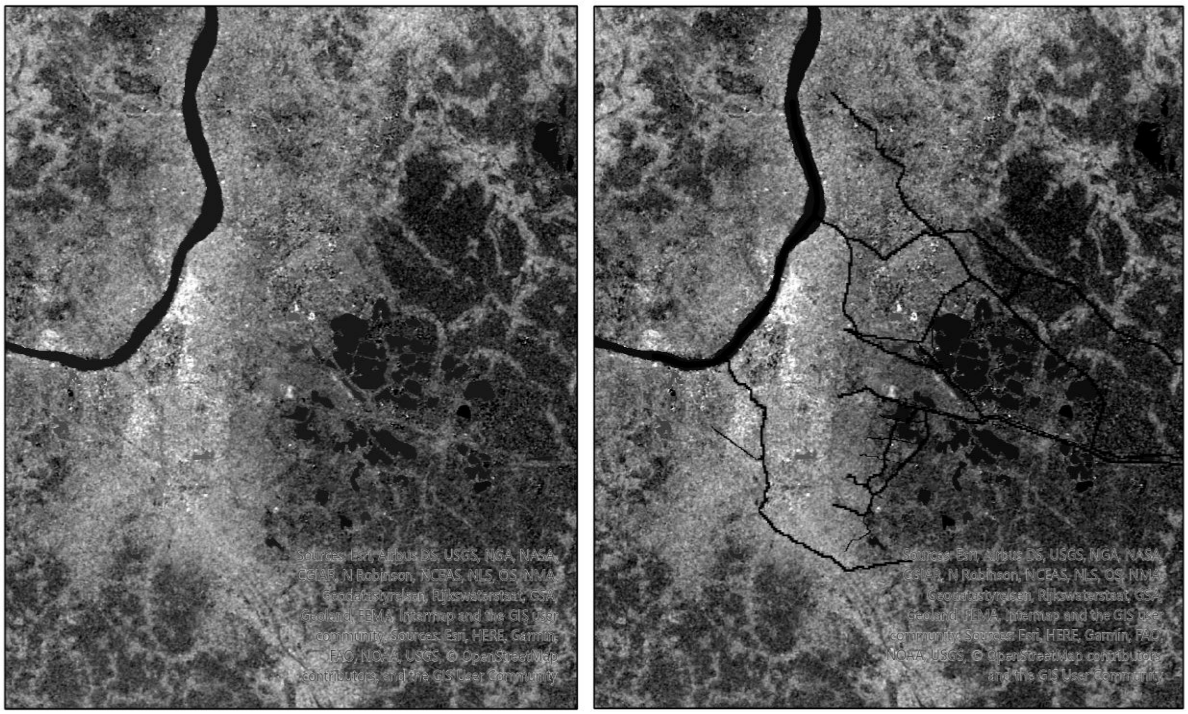


Figure A.5. Before DEM Reconditioning tool was applied (left) and after DEM Reconditioning tool was applied. The black lines burned into the DEM in the reconditioned map represent natural and artificial waterways (i.e., man-made canals) that were not captured through satellite imagery.

5. Terrain Processing

5.1. The reconditioned DEM was then input into the “Fill Sinks” tool to massage out the erroneous sinks and depressions.

5.2. The resulting output was used as the input in the “Flow Direction” tool. This function used an eight-direction flow model to determine which direction water would flow out of a pixel (i.e., toward the neighboring cell with the lowest elevation)[26]. The result was a raster of values that each represent a direction (1=East, 2=Southeast, 4=South, 8=Southwest, 16=West, 32=Northwest, and 64=North).

5.3. The output from the previous step was processed in the "Flow Accumulation" tool to sum the numbers of cells that flow into each pixel. Pixels with relatively high sums are areas where water flow concentrates based on the elevation profile to form streams and channels[26].

5.4. The "Raster Calculator" tool was used to define streams that had flow accumulation > 1,000 cells. This threshold value was selected because the output most closely resembled the stream network (Figure A.6). The "Stream Definition" tool can also be used for this step, and both tools are well documented throughout the literature.

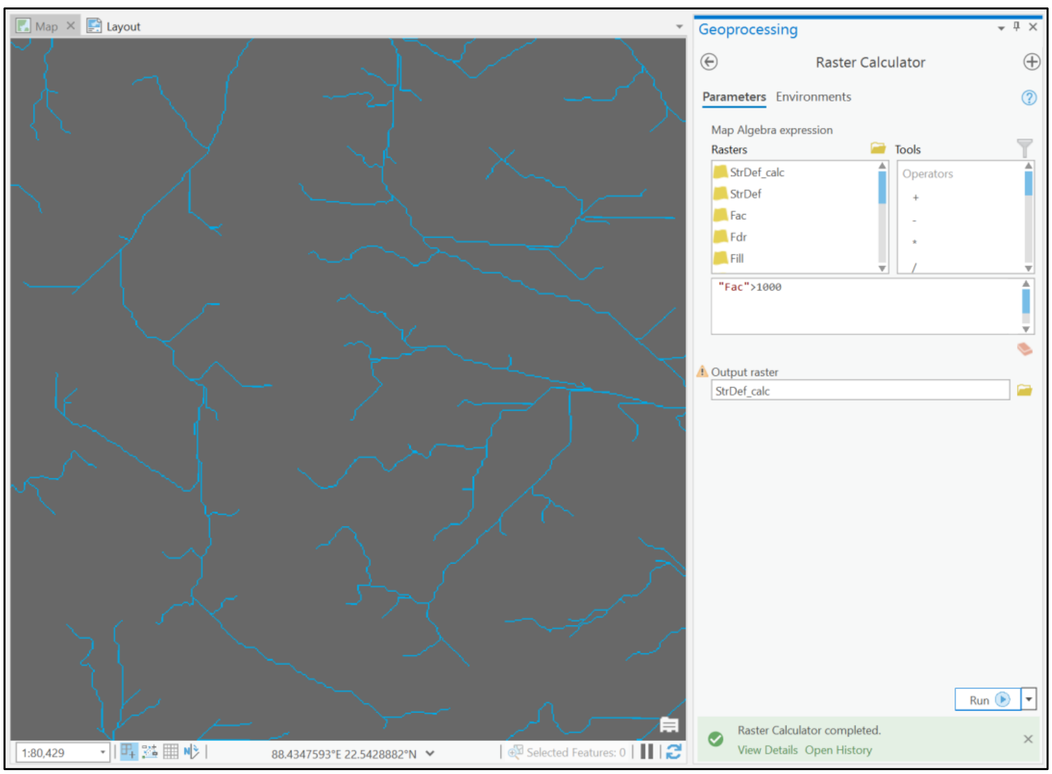


Figure A.6. Output from the "Raster Calculator" operation.

5.5. The “Stream Segmentation” tool was then used by inputting the stream raster from the previous step and the flow direction raster to assign a unique ID to each stream segment (Figure A.7).

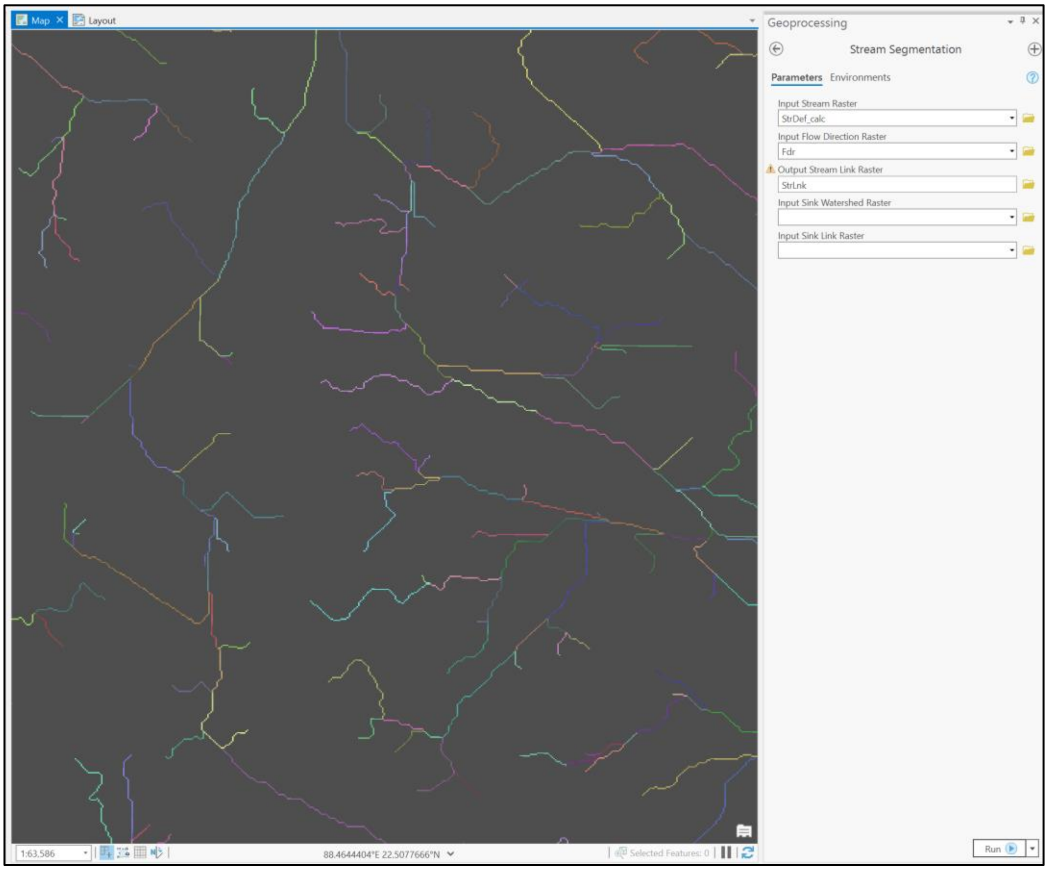


Figure A.7. Output from the “Stream Segmentation” operation.

5.6. The “Catchment Grid Delineation” tool was executed to separate the catchment areas and store the information into a grid. The stream segment raster generated in the previous step was input in addition to the flow direction raster to generate the catchment grid. The resulting delineated catchment areas are shown in Figure A.8.

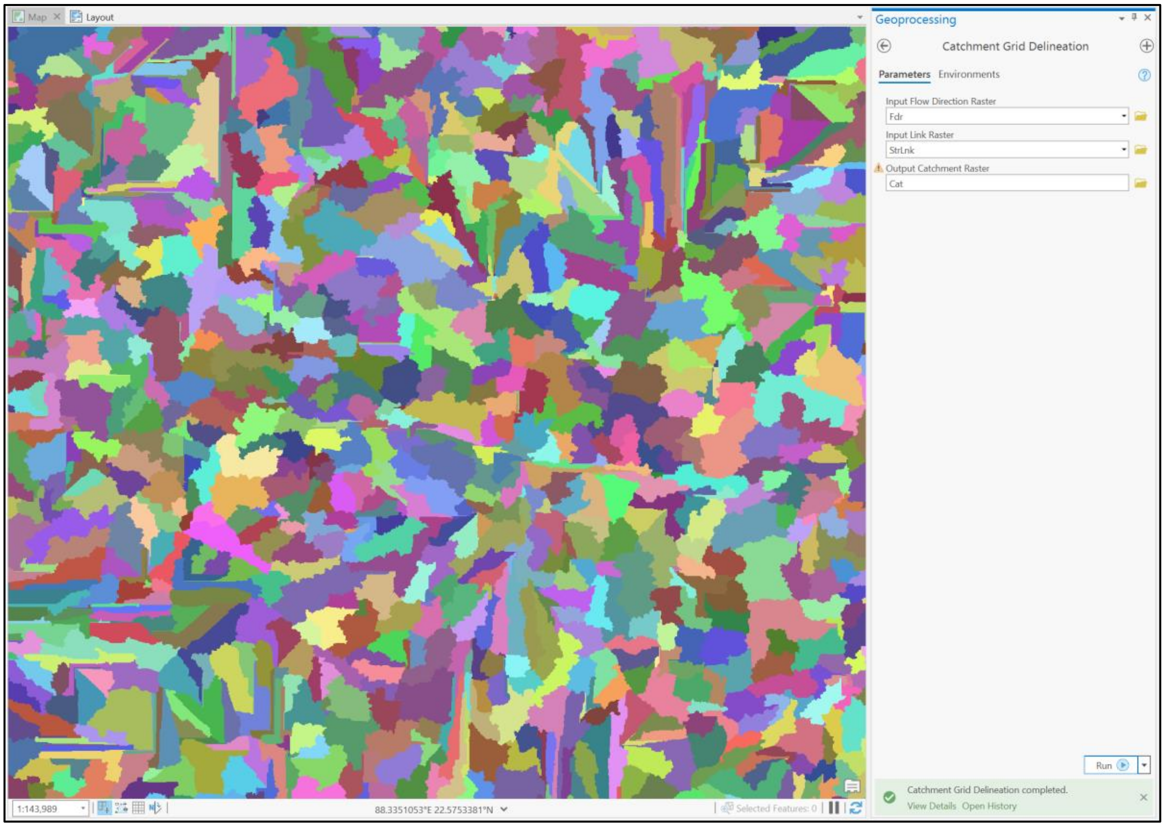


Figure A.8. Output from the “Catchment Grid Delineation” operation.

5.7. The “Catchment Polygon Processing” tool was then used to convert the catchment delineation grid into a polygon feature class.

6. Incorporating Sanitation Infrastructure

Diagrams that illustrate underground sewer and stormwater conveyance pipes were found online for three drainage systems in Kolkata: Town, Suburban, and Maniktala[44]. These were used to manually draw the underground lines that connect to open canals. Locations of pumping stations were previously collected through surveys prior to the COVID-19 pandemic, and these locations were added to the map. Figure A.9 depicts the connections between underground pipes, waterways, and pumping stations that were mapped in in ArcGIS.

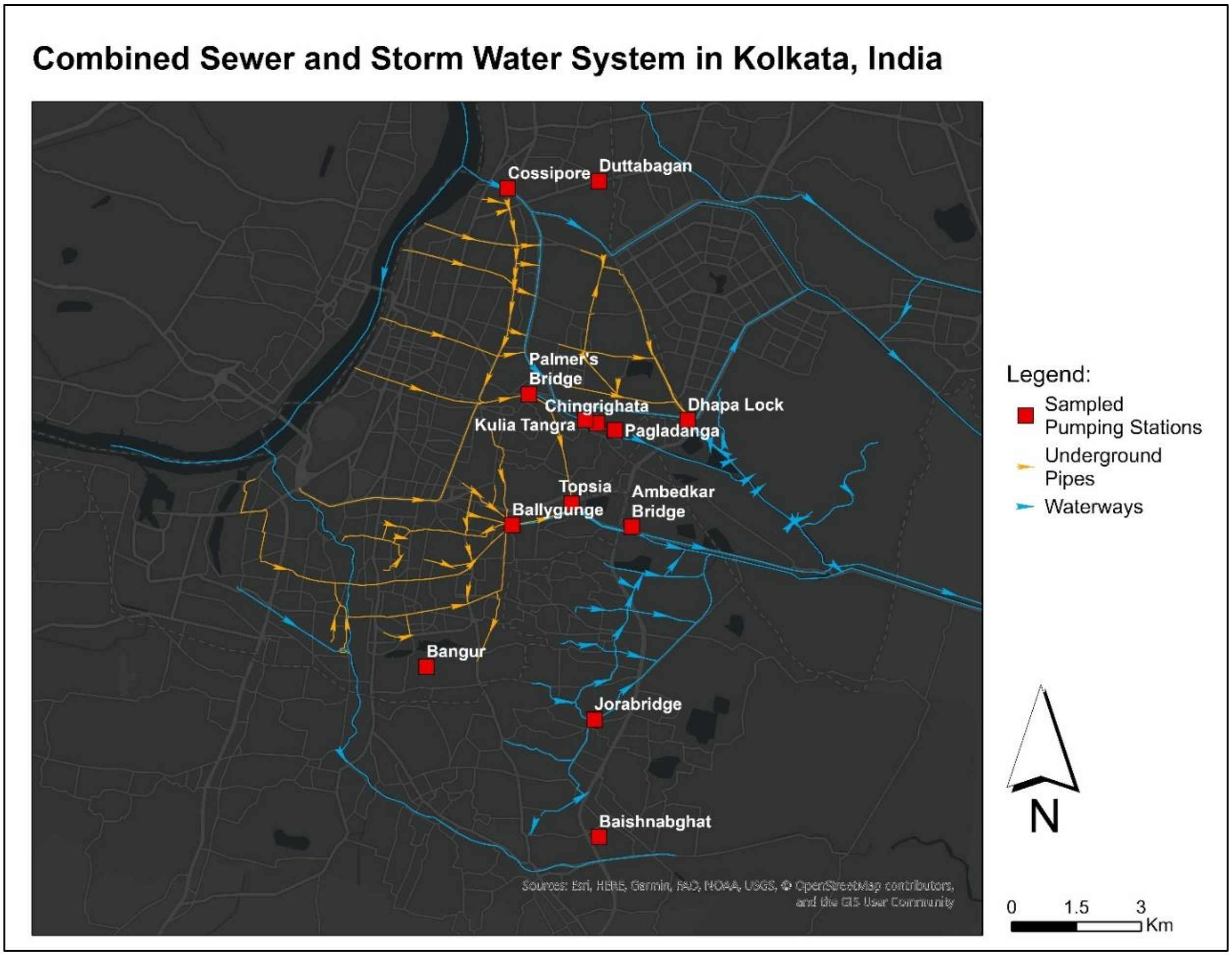


Figure A.9. Underground sewer and storm water pipes connect to open canals and waterways in the combined storm and sewer system in Kolkata, India. Pumping stations strategically move wastewater from areas of low- to high- elevation throughout the city.

7. Downloading and Overlaying Population Data

A shapefile with population data from the India Census of 2011 organized by wards was found within the ArcGIS online gallery[57]. Figure A.10 shows population categorization of wards using the Natural Breaks (Jenks) classification.

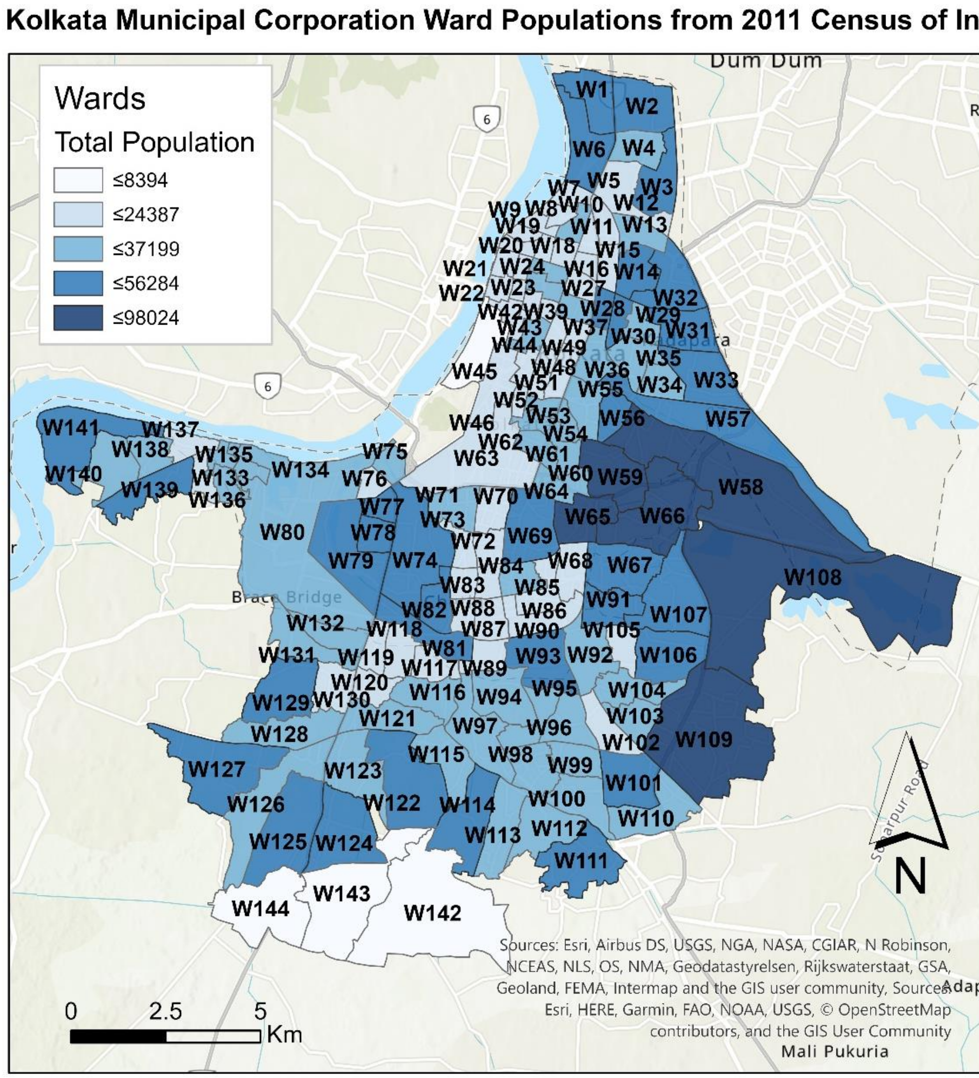


Figure A.10. Total population per ward from 2011 Census of India [42] using Natural Breaks (Jenks) classification

8. Estimating Catchment Areas of Sampled Pumping Stations

A process flow diagram was created to illustrate the decision-making process behind the population estimates for Kolkata and is shown in Figure A.11. This diagram is unique to Kolkata because not all urban settings will have a similar knowledge bank to draw from. A more generalized approach to estimating catchment area population size is presented for the second case study in Accra, Ghana in the Methods section of the thesis body.

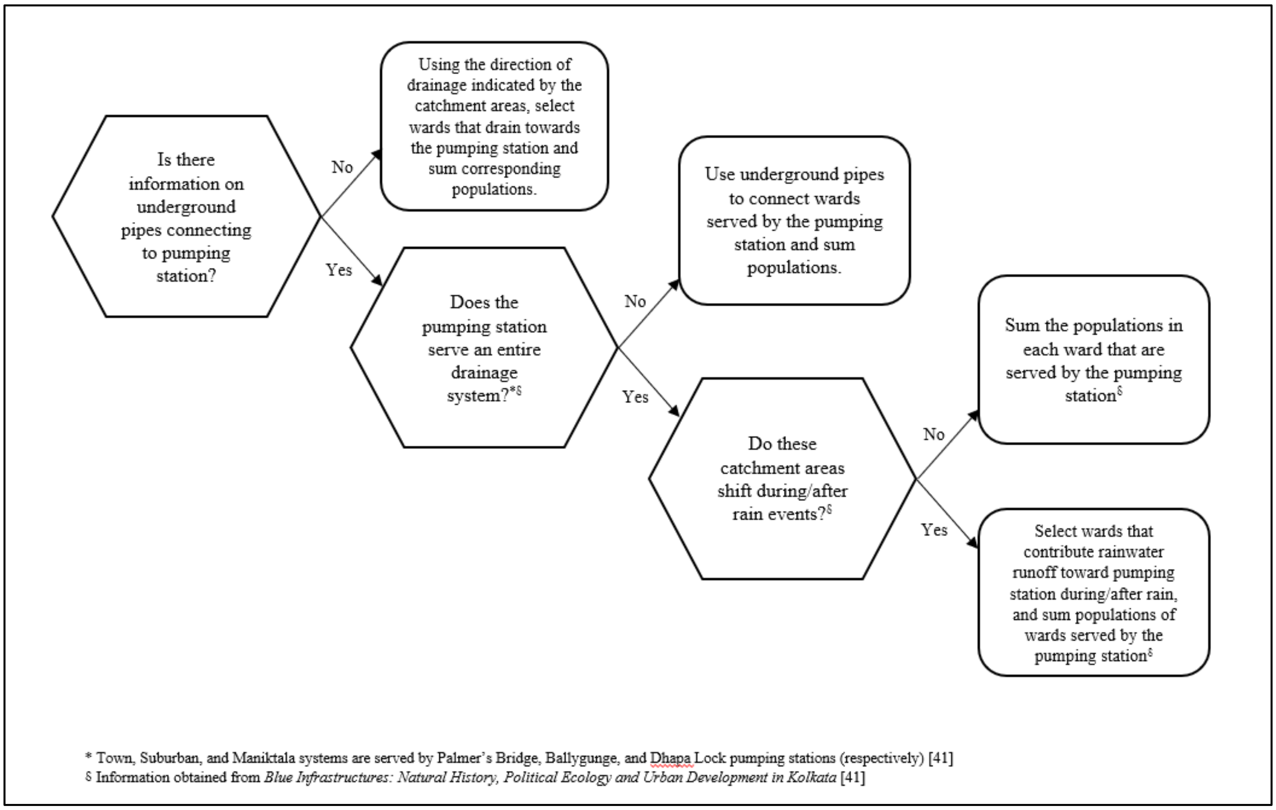


Figure A.11. Process Flow Diagram for Population Estimates of Sampled Pumping Stations in Kolkata, India

The population estimate procedure for the Duttabagan pumping station (PS) will be discussed in detail to demonstrate the decision-making process behind the calculations. The Duttabagan PS does not have information regarding the underground pipe network that connects sewer and storm water to it, and it is therefore in the first "No" box in the process flow diagram. It is worth noting that most of the sampled pumping stations met this criterion because there were only underground sewer diagrams for the Town, Suburban, and Maniktala systems[44]. The justification and rationale behind population estimates for all of the sampled pumping stations is provided in Appendix B.

The delineated catchment grid shown in Figure A.12a suggests that the catchment area that houses the Duttabagan PS drains southward. A zone above the facility was drawn to capture the upstream catchment area, and it reveals that partial areas of Wards 2, 3, and 4 drain to the Duttabagan PS (Figure A.12b).

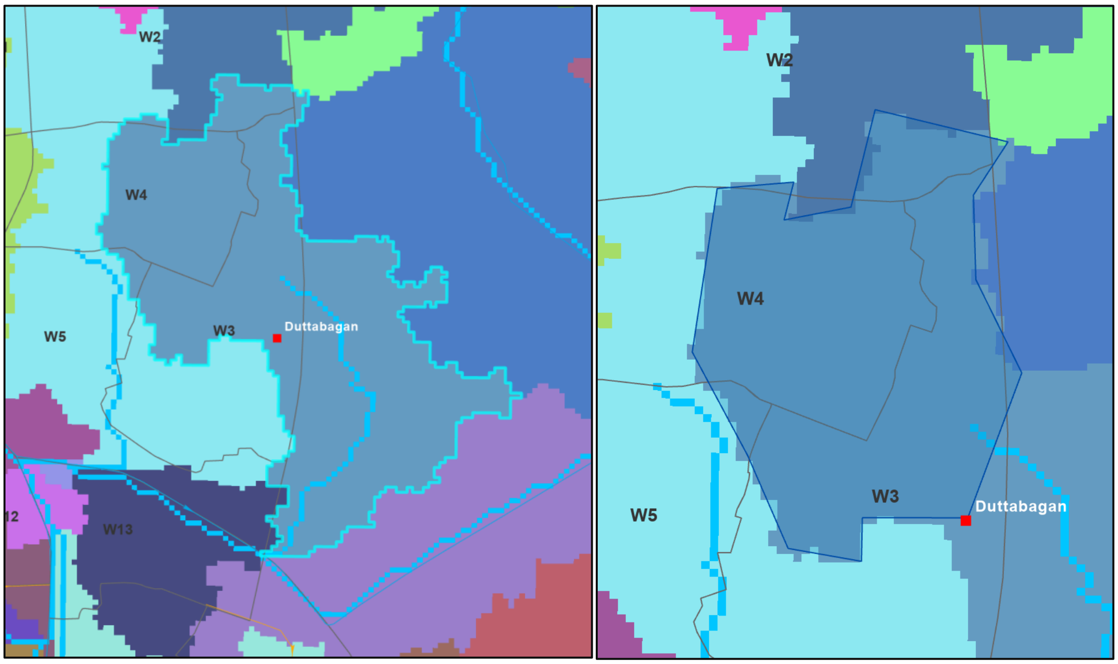


Figure A.12a (left). Primary catchment area in which Duttabagan PS is situated within. **Figure A.12b** (right). Zone of catchment area upstream of the Duttabagan PS.

The “Tabulate Intersection” tool was then applied to determine the percentage of area that each ward shares with the upstream catchment area (i.e., the drawn polygon). The inputs into this tool are shown in Figure A.13, and the output table results are shown in Figure A.14.

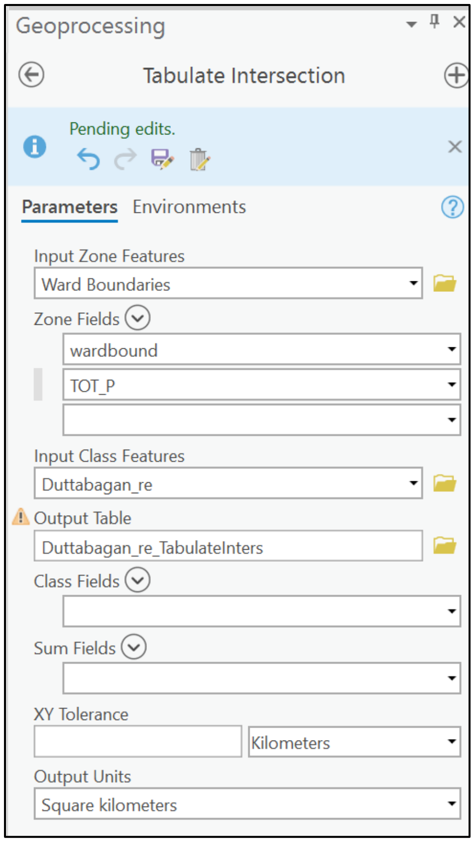
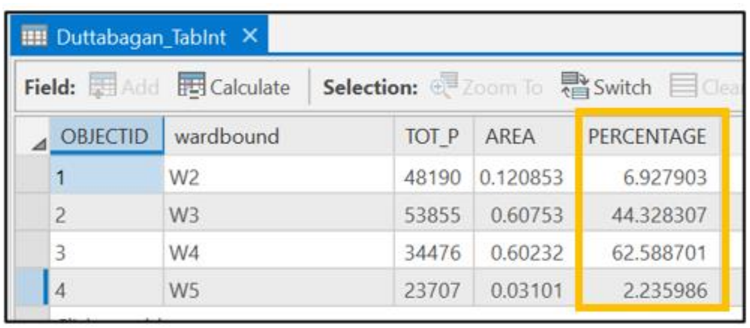


Figure A.13. Inputs into "Tabulate Intersection" tool.

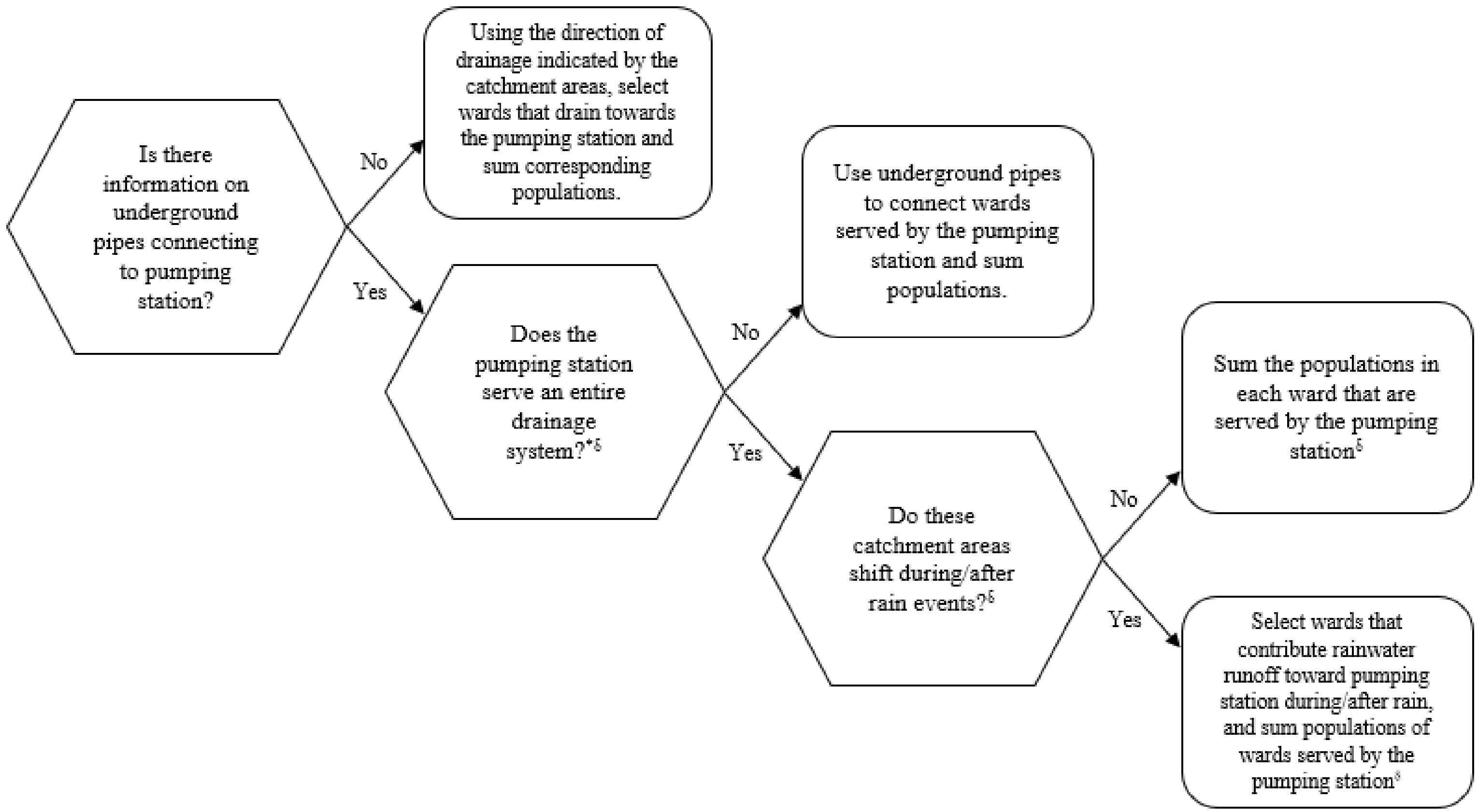


OBJECTID	wardbound	TOT_P	AREA	PERCENTAGE
1	W2	48190	0.120853	6.927903
2	W3	53855	0.60753	44.328307
3	W4	34476	0.60232	62.588701
4	W5	23707	0.03101	2.235986

Figure A.14. Screenshot of outputs from the "Tabulate Intersection" operation. The percentage column, outlined in orange, is used to approximate how much of the total population in each ward is served by the pumping station.

The percentages were multiplied by the total population in each corresponding ward and summed to calculate the total population estimate. Therefore, the population size contributing to the Duttabagan PS sample location is estimated to be 49,788 people.

APPENDIX B: Population Estimated of All Sampled Pumping Stations for Environmental Surveillance in Kolkata, India



* Town, Suburban, and Maniktala systems are served by Palmer's Bridge, Ballygunge, and Dhapa Lock pumping stations (respectively) [41]
§ Information obtained from *Blue Infrastructures: Natural History, Political Ecology and Urban Development in Kolkata* [41]

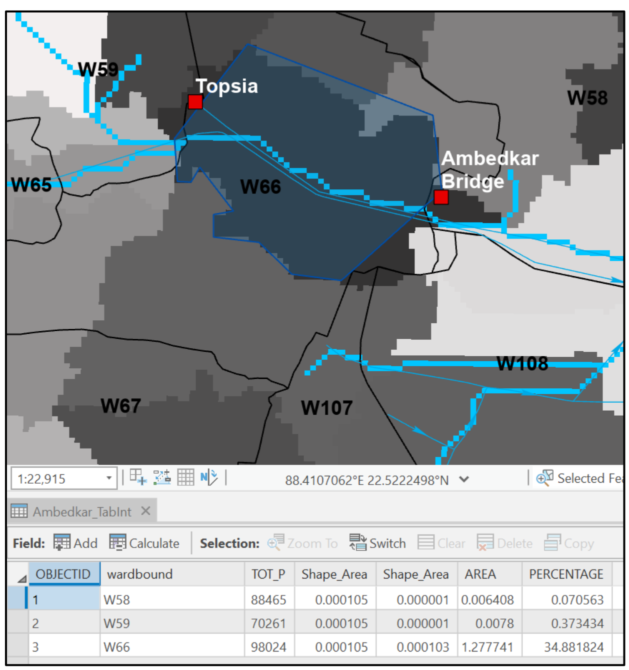
Figure B.1 Process Flow Diagram for Population Estimates of Sampled Pumping Stations in Kolkata, India

Ambedkar Bridge Pumping Station

Step in the Process Flow Diagram:

Using the direction of drainage indicated by the catchment areas, select wards that drain towards the pumping station and sum corresponding populations.

The Ambedkar Bridge PS does not have information regarding underground sewer and storm water lines that connect directly to the facility. Therefore, the catchment delineation grid was used to guide population estimate calculations. A zone was drawn to capture the upstream catchment area of the Ambedkar Bridge PS (Figure B.2). The "Tabulate Intersection" tool was then used to determine the percentage of area that each ward shares with the upstream catchment zone. These percentages were multiplied by the total population in each corresponding ward and summed to calculate the total population estimate. Therefore, the population size contributing to the Ambedkar Bridge PS sample location is estimated to be 34,517 people.



OBJECTID	wardbound	TOT_P	Shape_Area	Shape_Area	AREA	PERCENTAGE
1	W58	88465	0.000105	0.000001	0.006408	0.070563
2	W59	70261	0.000105	0.000001	0.0078	0.373434
3	W66	98024	0.000105	0.000103	1.277741	34.881824

Figure B.2 Upstream catchment zone of Ambedkar Bridge PS.

Baishnabghat Pumping Station

Step in the Process Flow Diagram:

Using the direction of drainage indicated by the catchment areas, select wards that drain towards the pumping station and sum corresponding populations.

The Baishnabghat PS does not have information regarding underground sewer and storm water lines that connect directly to the facility. Therefore, the catchment delineation grid was used to guide population estimate calculations. A zone was drawn to capture the upstream catchment area of the facility (Figure B.3). It was noted that this zone contains area outside of the ward boundaries, and therefore the final estimate is not counting people who live in this area.

The "Tabulate Intersection" tool was then used to determine the percentage of area that each ward shares with the upstream catchment zone. These percentages were multiplied by the total population in each corresponding ward and summed to calculate the total population estimate. Therefore, the population size contributing to the Baishnabghat PS sample location is estimated to be 5,916 people.

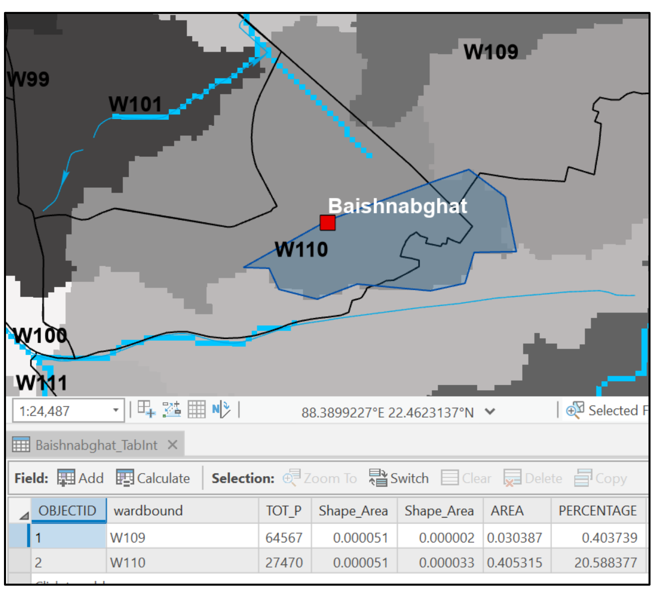


Figure B.3 Upstream catchment zone of Baishnabghat Pumping Station.

Ballygunge Pumping Station

Step in the Process Flow Diagram:

Select wards that contribute rainwater runoff toward pumping station during/after rain, and sum populations served by the pumping station.[§]

Ballygunge serves the Suburban System and is the largest pumping station in the KMC[43–45]. There is information about underground pipe connections leading to this facility; however, these connections were particularly difficult to interpret in the "Suburban System Sewer Network" diagram due to disjointed lines, lack of directionality indicated for some sewers, and an incomplete legend that specifies the symbology used. Additionally, these diagrams do not completely agree with data from KMC (2005) regarding the wards fully and partially covered within the Suburban System[45]. As Figure B.4 shows, the partially covered wards are distributed randomly amongst the fully covered wards in an unintelligible pattern. It is not possible from this information to surmise how much of each ward is actually covered under the Suburban System and sent through the Ballygunge PS. For this reason, the catchment delineation grid will be used to help guide the population size calculation for partial wards. The external sources of information will be consulted and applied through logical reasoning.

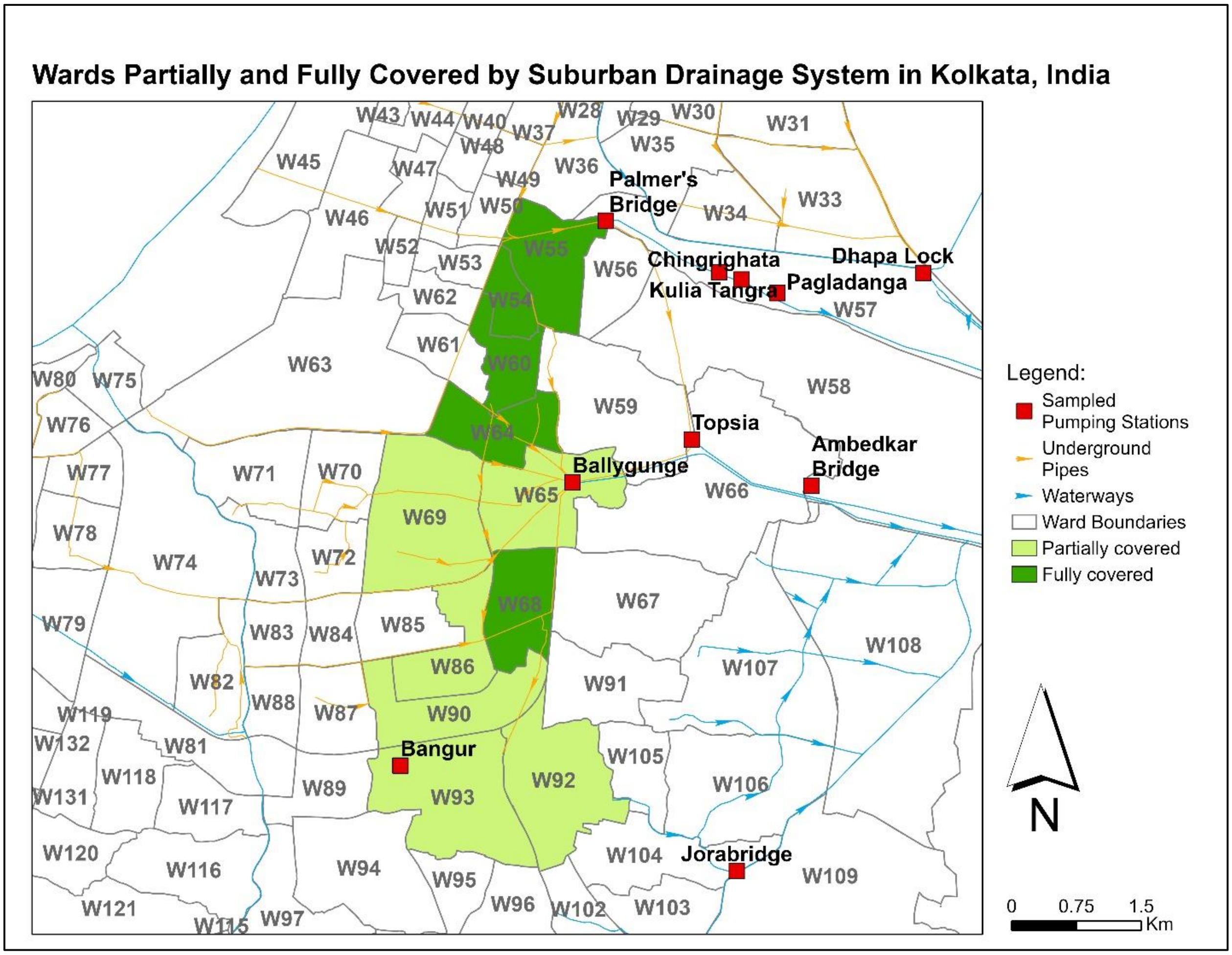


Figure B.4 According to KMC (2005), these wards are partially and fully covered by the Suburban System, which discharges to the Ballygunge PS.

A zone was drawn over catchment areas that intersected with the partially covered wards and drained towards the Ballygunge PS (Figure B.5). The "Tabulate Intersection" tool was then used to determine the percentage of area the wards shared with the zone, and these percentages were multiplied by the corresponding total population in each ward. These calculations were added to the population numbers from fully covered wards. Therefore, the population size contributing to the Ballygunge PS is estimated to be 225,074 people.

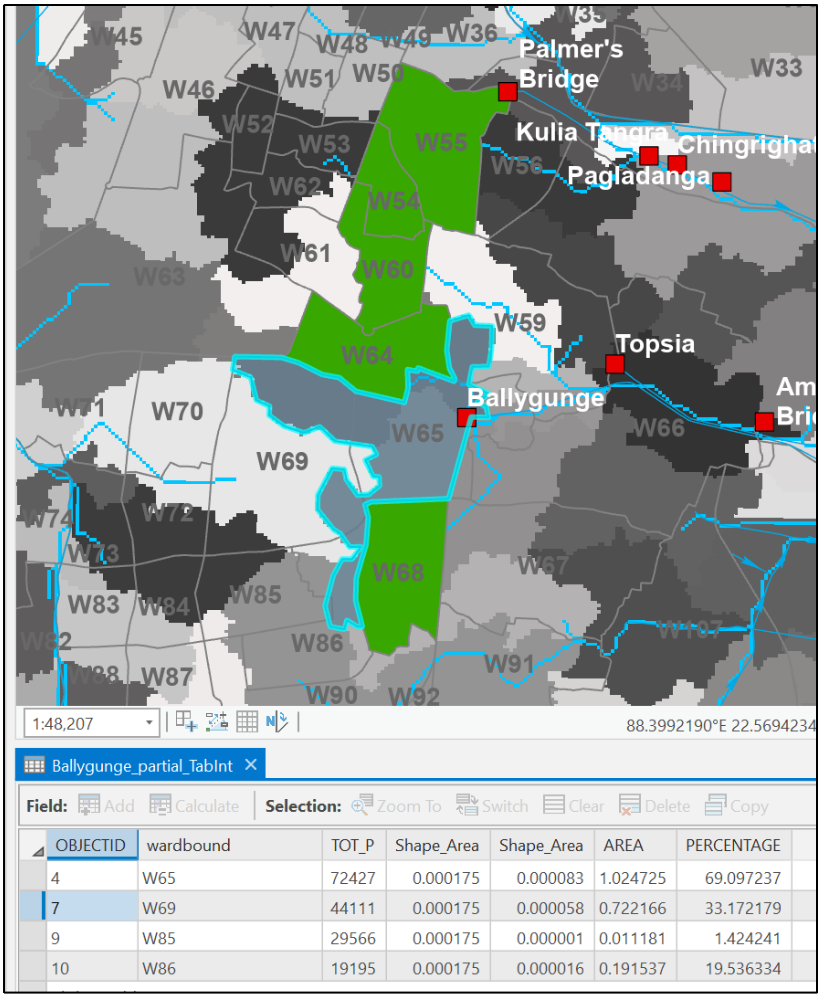


Figure B.5 Upstream catchment zone of Ballygunge PS for wards partially covered in the Suburban System.

Bangur Pumping Station

Step in the Process Flow Diagram:

Using the direction of drainage indicated by the catchment areas, select wards that drain towards the pumping station and sum corresponding populations.

The Bangur PS does not have information regarding underground sewer and storm water lines that connect directly to the facility. Therefore, the catchment delineation grid was used to guide population estimate calculations. The Bangur PS is located near the watershed boundary line where two catchment areas flow away from one another. Therefore, the zone drawn to capture the upstream catchment area of the facility is relatively small (Figure B.6). Unfortunately, the Irrigation and Waterways Department does not have Bangur PS listed on its website to cross-reference the installed capacity[43].

The "Tabulate Intersection" tool was then used to determine the percentage of area that each ward shares with the upstream catchment zone of the Bangur PS. These percentages were multiplied by the total population in each corresponding ward and summed to calculate the total population estimate. Therefore, the population size contributing to the Bangur PS sample location is estimated to be 1,525 people.

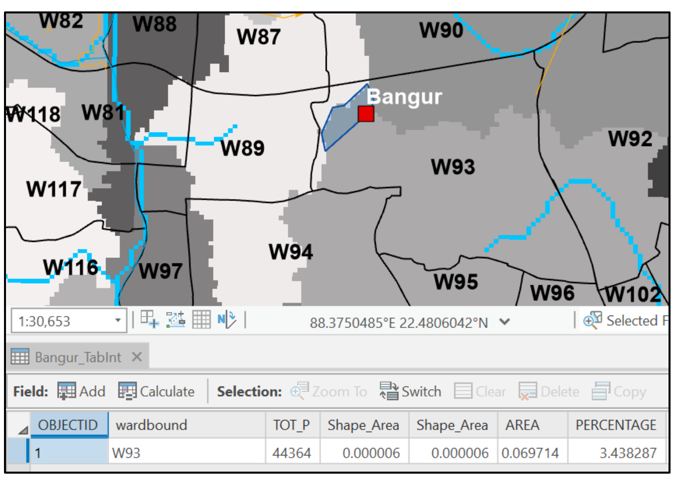


Figure B.6. Upstream catchment zone of Bangur PS.

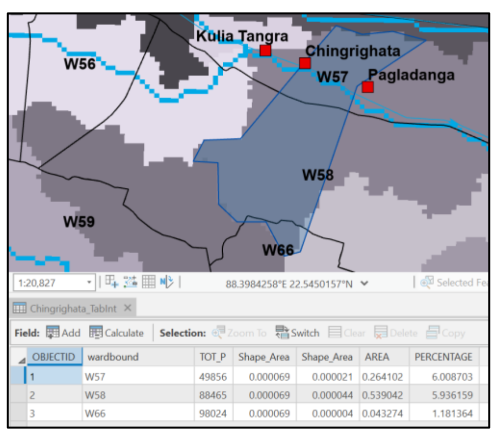
Chingrighata Pumping Station

Step in the Process Flow Diagram:

> Using the direction of drainage indicated by the catchment areas, select wards that drain towards the pumping station and sum corresponding populations.

The Chingrighata PS does not have information regarding underground sewer and storm water lines that connect directly to the facility. This facility is situated along the Town Head Cut (THC) Channel that the Palmer's Bridge PS discharges into, and it is downstream of the Kulia Tangra PS (Figure B.7a). The capacity of Chingrighata PS (100 cusecs) is well below the capacity of Palmer's Bridge PS (1,184 cusecs)[43]. With this knowledge and drawing from the catchment delineation map, it can be surmised that the facility does not receive water from the THC Channel, but rather that it discharges wastewater into it.

A zone was drawn to capture the upstream catchment area of Chingrighata PS (Figure B.7b). The "Tabulate Intersection" tool was then used to determine the percentage of area that each ward shares with the upstream catchment area. These percentages were multiplied by the total population in each corresponding ward and summed to calculate the total population estimate. Therefore, the population size contributing to the Chingrighata PS sample location is estimated to be 9,405 people.



OBJECTID	wardbound	TOT_P	Shape_Area	Shape_Area	AREA	PERCENTAGE
1	W57	49856	0.000069	0.000021	0.264102	6.008703
2	W58	88465	0.000069	0.000044	0.539042	5.936159
3	W66	98024	0.000069	0.000004	0.043274	1.181364

Figure B.7a (left). Location of Chingrighata PS relative to other stations and canals. **Figure B.7b** (right). Upstream catchment zone of Chingrighata PS.

Cossipore Pumping Station

Step in the Process Flow Diagram:

> Using the direction of drainage indicated by the catchment areas, select wards that drain towards the pumping station and sum corresponding populations.

Although the Cossipore PS does not have information regarding how sewer and storm water lines connect to the facility, it is adjacent to the Town System, which is entirely served by the Palmer's Bridge PS. Factoring in this proximal information, it appears that Wards 1 and 6 (north of the facility) are most likely served by the Cossipore PS (Figure B.8). Unfortunately, the Irrigation and Waterways Department does not have Bangur PS listed on its website to cross-reference the installed capacity[43]. However, due to a lack of competing pumping stations in the area, it is reasonable to assume that entire populations of Wards 1 and 6 are served by this PS. Therefore, the population size for the Cossipore PS is estimated to be 95,471 people.

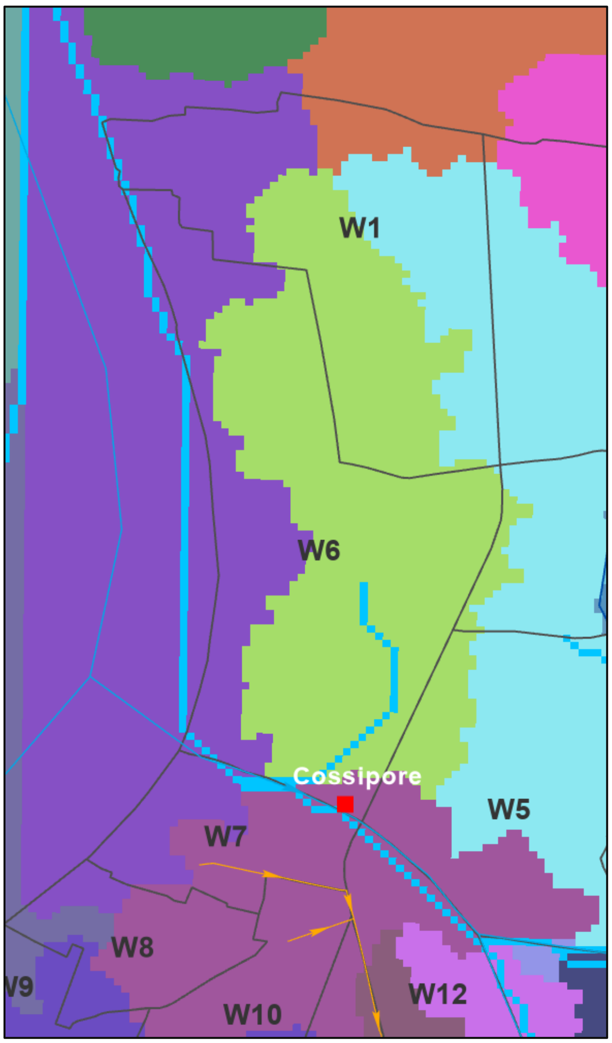
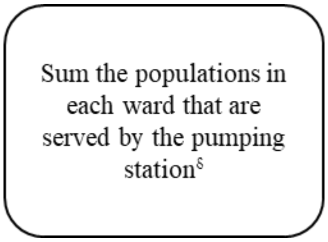


Figure B.8 Wards 1 and 6 were assumed to drain to the Cossipore Pumping Station due to lack of competiting pumping stations in the area.

Dhapa Lock Pumping Station

Step in the Process Flow Diagram:



The Dhapa Lock PS is one of the largest pumping stations in the KMC, and it serves the Maniktala System[44,45]. Sewer diagrams were available for this system, so it was possible to trace underground pipes to the Dhapa Lock PS[44]. KMC (2005) indicates that several wards are fully served (wards 13-14 and 29-35) by the Maniktala System and that Ward 5 is partially served by this system (Figure B.9)[45].

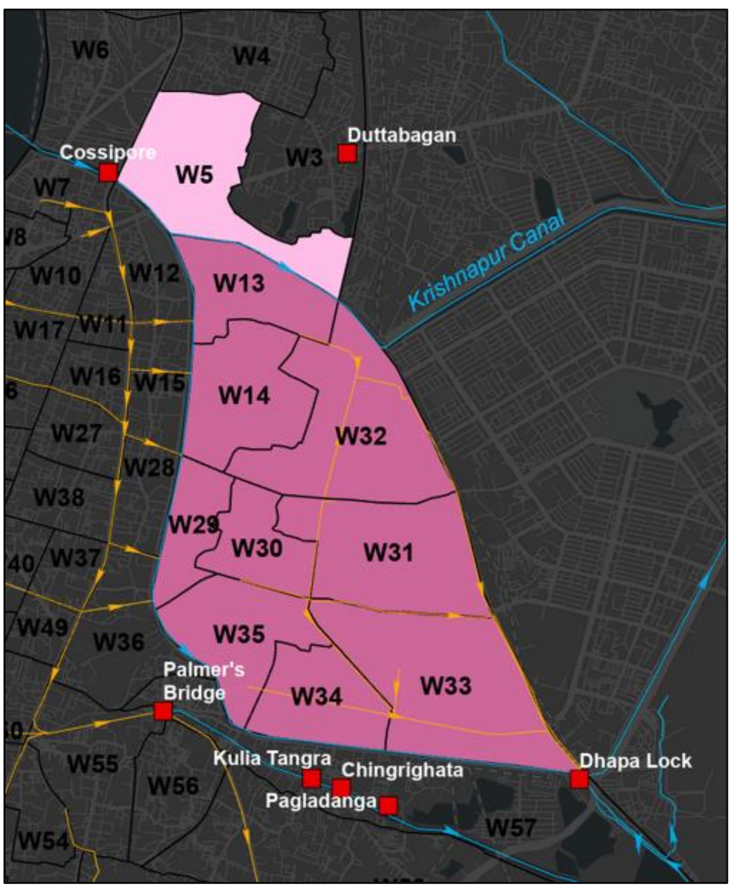
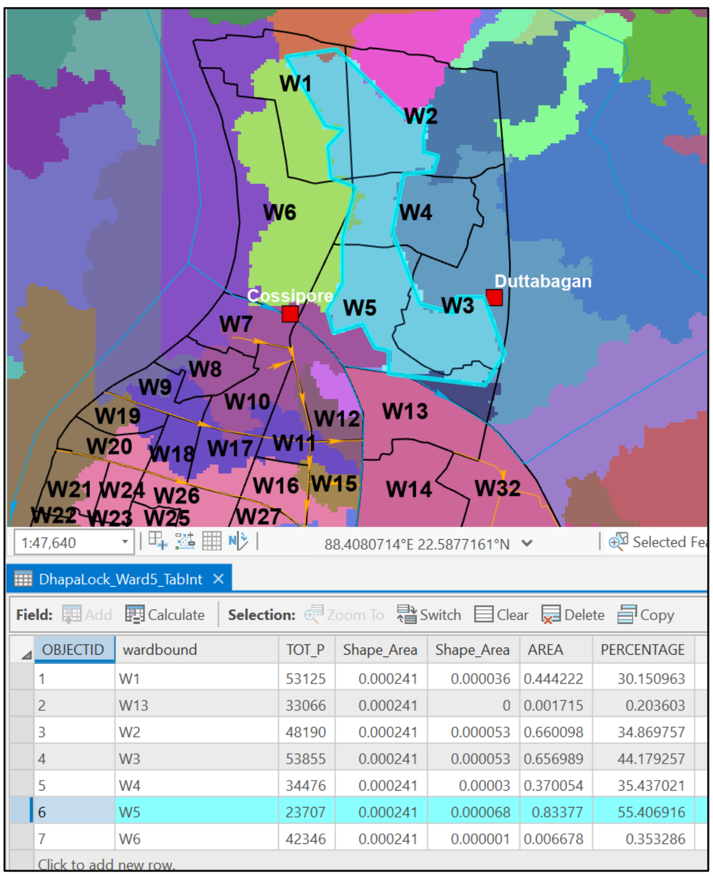


Figure B.9 Wards partially (light pint) and fully covered (dark pink) by the Maniktala System, which drains to the Dhapa Lock Pumping Station.

Building off this knowledge, the populations of the fully covered wards were summed. To account for the partial service of Ward 5, the catchment grid layer was used to estimate the percentage of area from Ward 5 that may be contributing to the Dhapa Lock PS through unknown underground pipes. It was noted that rainwater runoff that does not find its way into underground storm water lines will drop into the Krishnapur (or Kestopur) Canal, and therefore it will not flow overland to the Dhapa Lock PS.

A zone was drawn to capture the upstream catchment area in common with Ward 5 (Figure B.10). The “Tabulate Intersection” tool was then used to determine the percentage of area, which was then multiplied by the total population of Ward 5. Adding this to the population numbers from fully covered wards, the population size contributing to the Dhapa Lock PS sample location is estimated to be 378,753 people.



OBJECTID	wardbound	TOT_P	Shape_Area	Shape_Area	AREA	PERCENTAGE
1	W1	53125	0.000241	0.000036	0.444222	30.150963
2	W13	33066	0.000241	0	0.001715	0.203603
3	W2	48190	0.000241	0.000053	0.660098	34.869757
4	W3	53855	0.000241	0.000053	0.656989	44.179257
5	W4	34476	0.000241	0.00003	0.370054	35.437021
6	W5	23707	0.000241	0.000068	0.83377	55.406916
7	W6	42346	0.000241	0.000001	0.006678	0.353286

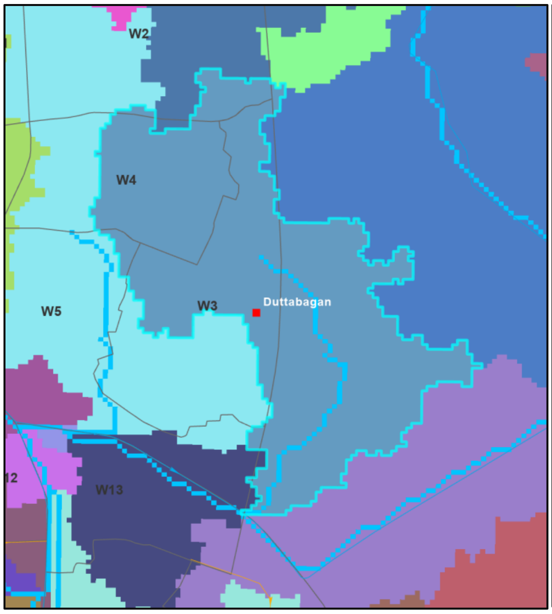
Figure B.10 Upstream catchment area of Maniktala System to estimate how much of Ward 5 potentially drains to the Dhapa Lock Pumping Station.

Duttabagan Pumping Station

Step in the Process Flow Diagram:

> Using the direction of drainage indicated by the catchment areas, select wards that drain towards the pumping station and sum corresponding populations.

The was no information available about the underground pipe network that connects sewer and storm water to the Duttabagan PS, and it is therefore in the first “No” box in the process flow diagram. The delineated catchment grid shown in Figure B.11a suggests that the catchment area that houses the Duttabagan PS drains southward. A polygon was drawn to represent the smaller catchment area that drains toward the facility (Figure B.11b). This process revealed that partial areas of Wards 2, 3, and 4 drain to the Duttabagan PS. The “Tabulate Intersection” tool was then applied to determine the percentage of area that Wards 2, 3, and 4 overlap with the drawn polygon, and the output is shown in Figure B.12. The percentages were multiplied by the total population in each ward and summed to calculate the total population size of the catchment area.



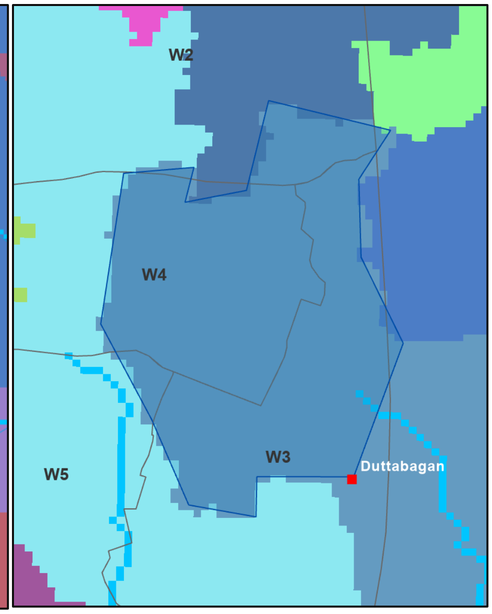


Figure B.11a (left). Watershed in which Duttabagan PS is situated. **Figure B.11b** (right). Perceived zone of catchment area that drains to the Duttabagan PS.

OBJECTID	wardbound	TOT_P	Shape_Area	Shape_Area	AREA	PERCENTAGE
1	W2	48190	0.000122	0.000011	0.131143	6.927617
2	W3	53855	0.000122	0.000054	0.672131	45.19749
3	W4	34476	0.000122	0.000053	0.653586	62.588479
4	W5	23707	0.000122	0.000003	0.033649	2.236103

Figure B.12 Screenshot of outputs from the "Tabulate Intersection" operation. The percentage column is used to approximate how much of the total population in each ward is served by the Duttabagan PS.

The percentages were multiplied by the total population in each corresponding ward and summed to calculate the total population estimate. Therefore, the population size contributing to the Duttabagan PS sample location is estimated to be 49,788 people.

Jorabridge Pumping Station

Step in the Process Flow Diagram:

Using the direction of drainage indicated by the catchment areas, select wards that drain towards the pumping station and sum corresponding populations.

The Jorabrige PS does not have information about how underground sewer and storm water lines that connect directly to the facility, but it is one of the pumping stations situated within a robust canal system. The upstream catchment area was estimated using the catchment delineation grid with the assumption that much of the drainage in this area drops into the canals before reaching the Jorabridge PS (Figure B.13).

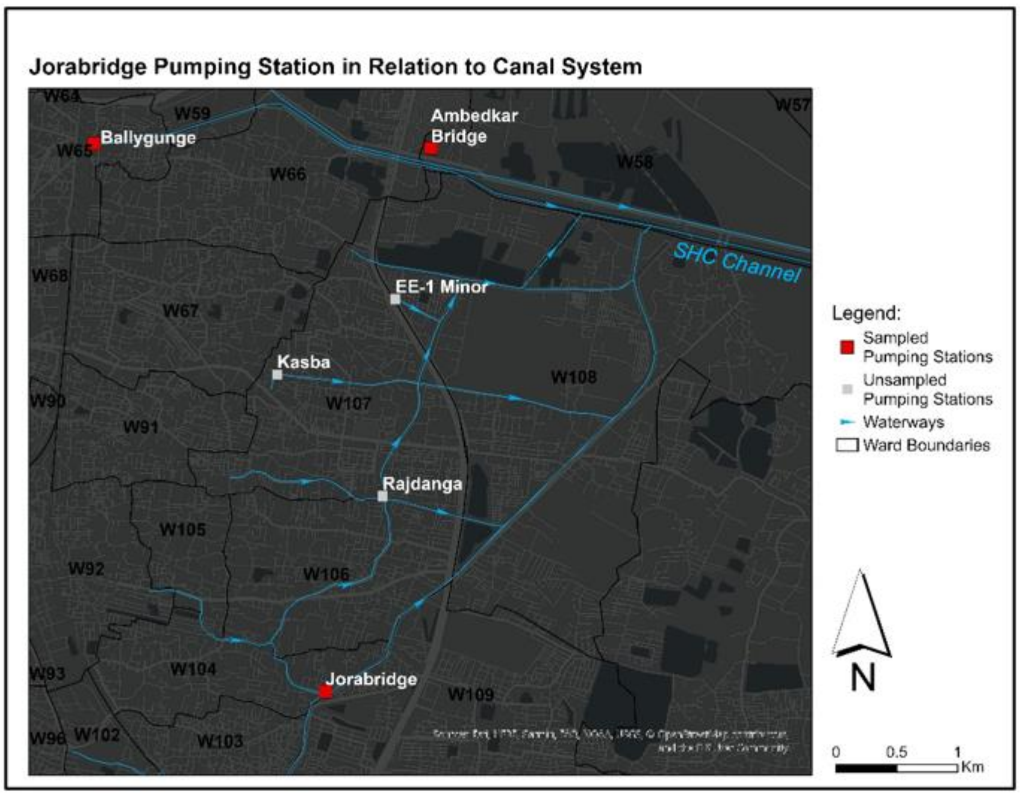
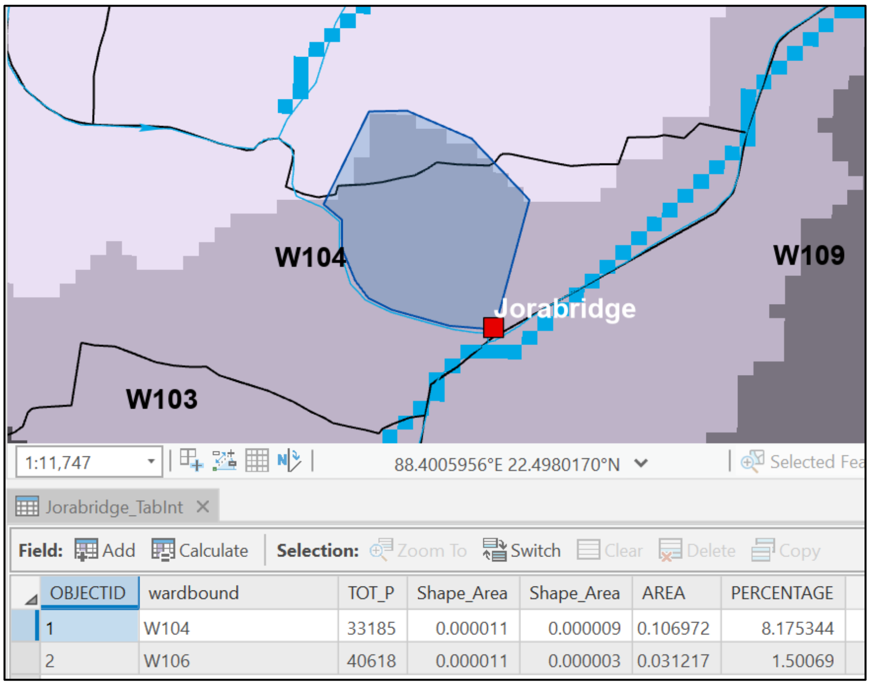


Figure B.13 The Jorabridge PS is situated within a robust network of open, man-made canals. It the southern-most pumping station in the network, which generally flows northeast toward the SHC Channel.

A zone was drawn to estimate the upstream catchment area leading to the Jorabridge PS (Figure B.14). The "Tabulate Intersection" tool was then used to determine the percentage of area that each ward shares with the upstream catchment area. These percentages were multiplied by the total population in each corresponding ward and summed to calculate the total population estimate. Therefore, the population size contributing to the Jorabridge PS sample location is estimated to be 3,323 people.



OBJECTID	wardbound	TOT_P	Shape_Area	Shape_Area	AREA	PERCENTAGE
1	W104	33185	0.000011	0.000009	0.106972	8.175344
2	W106	40618	0.000011	0.000003	0.031217	1.50069

Figure B.14 Upstream catchment area of the Jorabridge Pumping Station.

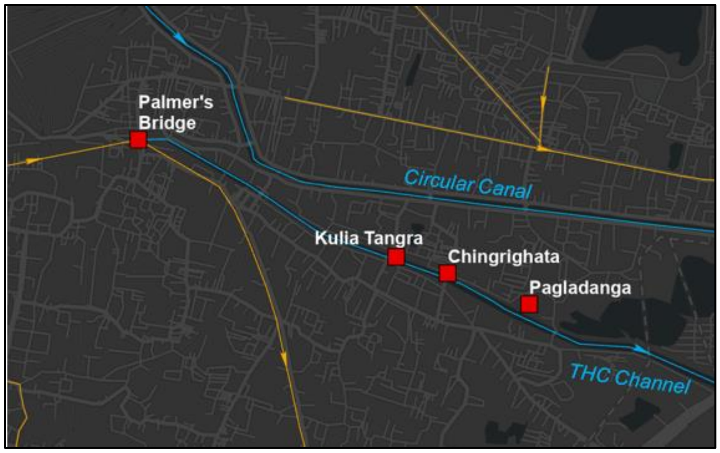
Kulia Tangra Pumping Station

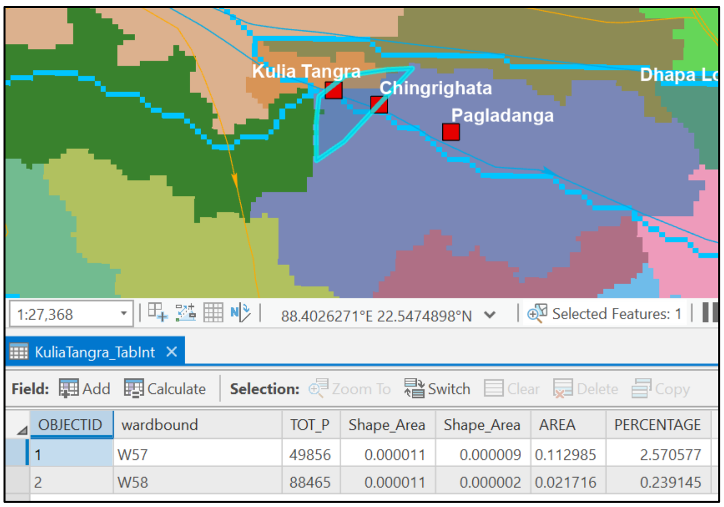
Step in Process Flow Diagram:

Using the direction of drainage indicated by the catchment areas, select wards that drain towards the pumping station and sum corresponding populations.

The Kulia Tangra PS does not have information regarding the sewer and storm water lines that connect directly to the facility, but it is situated between the Maniktala and Town systems. The proximal information helps inform which areas are draining to this location. From Figure B.15a, it can be seen that the Kulia Tangra PS is situated along the THC Channel that Palmer's Bridge PS discharges into. The capacity of Kulia Tangra (40 cusecs) is well below the capacity of Palmer's Bridge (1,184 cusecs)[43]. Similar to the Chingrighata PS, it can be surmised that the Kulia Tangra PS does not receive water from the THC Channel, but rather that it discharges wastewater into the channel.

A zone was drawn to capture the upstream catchment area of Kulia Tangra PS (Figure B.15b). The "Tabulate Intersection" tool was then used to determine the percentage of area that each ward shares with the upstream catchment area. These percentages were multiplied by the total population in each corresponding ward and summed to calculate the total population estimate. Therefore, the population size contributing to the Kulia Tangra PS sample location is estimated to be 1,493 people.





OBJECTID	wardbound	TOT_P	Shape_Area	Shape_Area	AREA	PERCENTAGE
1	W57	49856	0.000011	0.000009	0.112985	2.570577
2	W58	88465	0.000011	0.000002	0.021716	0.239145

Figure B.15a (top). Location of Kulia Tangra PS relative to other stations and canals. **Figure B.15b** (bottom). Upstream catchment zone of Kulia Tangra PS.

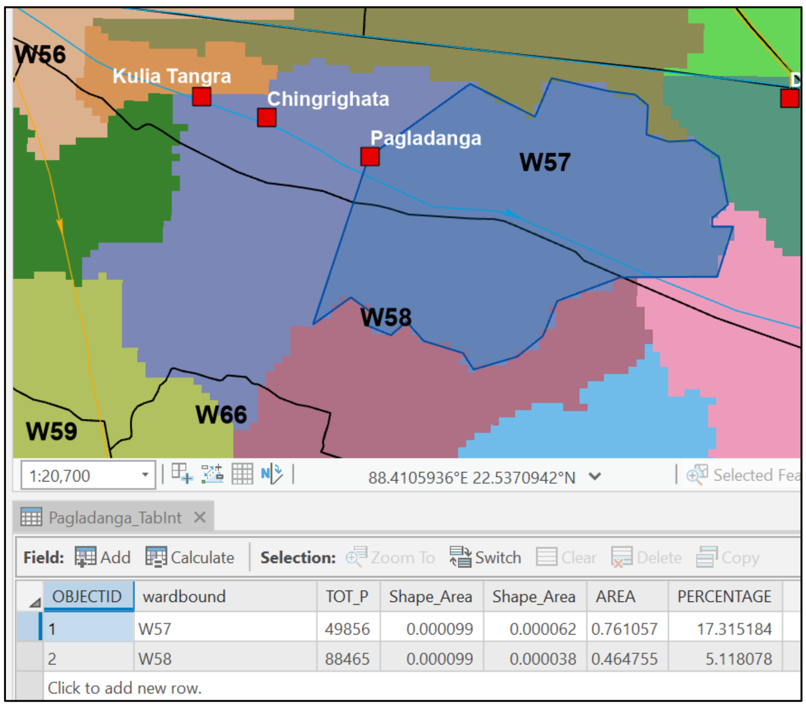
Pagladanga Pumping Station

Step in the Process Flow Diagram:

> Using the direction of drainage indicated by the catchment areas, select wards that drain towards the pumping station and sum corresponding populations.

The Pagladanga PS does not have information regarding the sewer and storm water lines that connect directly to the facility, but it is situated between the Maniktala and Town systems. The proximal information helps inform which areas are draining to this location. From Figure B.16a, it can be seen that the Pagladanga PS is situated along the THC Channel that Palmer's Bridge PS discharges into, and it is downstream of the Chingrighata and Kulia Tangra pumping stations. The capacity of Pagladanga PS (48 cusecs) is well below the capacity of Palmer's Bridge (1,184 cusecs)[43]. Similar to the Chingrighata and Kulia Tangra stations, it can be surmised that the Pagladanga PS does not receive water from the THC Channel, but rather that it discharges wastewater into the channel.

A zone was drawn to capture the upstream catchment area of Pagladanga PS (Figure B.16b). The "Tabulate Intersection" tool was then used to determine the percentage of area that each ward shares with the upstream catchment area. These percentages were multiplied by the total population in each corresponding ward and summed to calculate the total population estimate. Therefore, the population size contributing to the Pagladanga PS sample location is estimated to be 13,160 people.



OBJECTID	wardbound	TOT_P	Shape_Area	Shape_Area	AREA	PERCENTAGE
1	W57	49856	0.000099	0.000062	0.761057	17.315184
2	W58	88465	0.000099	0.000038	0.464755	5.118078

Figure B.16a (top). Location of Pagladanga PS relative to other stations and canals. **Figure B.16b** (bottom). Upstream catchment zone of Pagladanga PS.

Palmer's Bridge Pumping Station

Step in the Process Flow Diagram:

Sum the populations in each ward that are served by the pumping station[§]

The Palmer's Bridge PS is one of the largest pumping stations in the KMC, and combined sewer and storm water from the Town System is funneled to this facility[44,45]. Coherent integration of the various sources of information is difficult for Palmer's Bridge because there are inconsistencies in the data. In the book *Blue Infrastructures: Natural History, Political Ecology and Urban Development in Kolkata* by Jenia Mukherjee (2020), Table 5.4 (source: KMC, 2005) summarizes how wards are covered by major drainage systems[45]. The table specifies that Wards 6 and 63 are partially served by the Town System, however Map 5.1 shows Ward 63 in its entirety draining towards Palmer's Bridge[45]. Regarding Ward 6, it was determined from the catchment delineation grid that it most likely drains to the Cossipore PS, where wastewater then discharges into an open canal and bypasses the Palmer's Bridge PS (see Cossipore PS for more information). Therefore, it was assumed that Ward 6 does not contribute wastewater to the Palmer's Bridge PS, while Ward 63 was assumed to be fully covered by Palmer's Bridge PS due to lack of competing pumping stations in the area. The populations from wards fully covered by the Town System plus the total population from Ward 63 were summed (41 wards total) to estimate the population size (Figure 17). Therefore, the population size for the Palmer's Bridge PS is estimated to be 890,681 people.

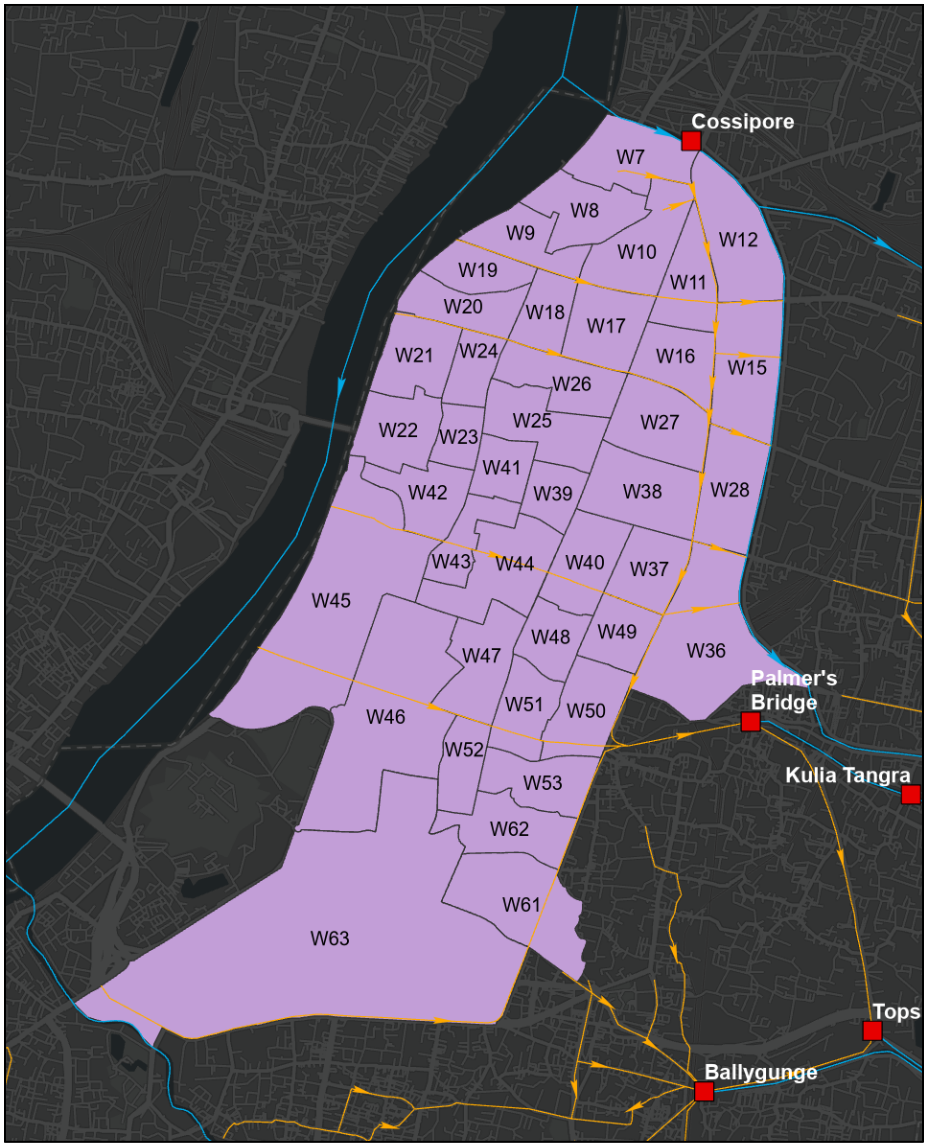


Figure B.17 Wards included in the population size estimate for the Palmer's Bridge PS.

Topsia (Old) Pumping Station

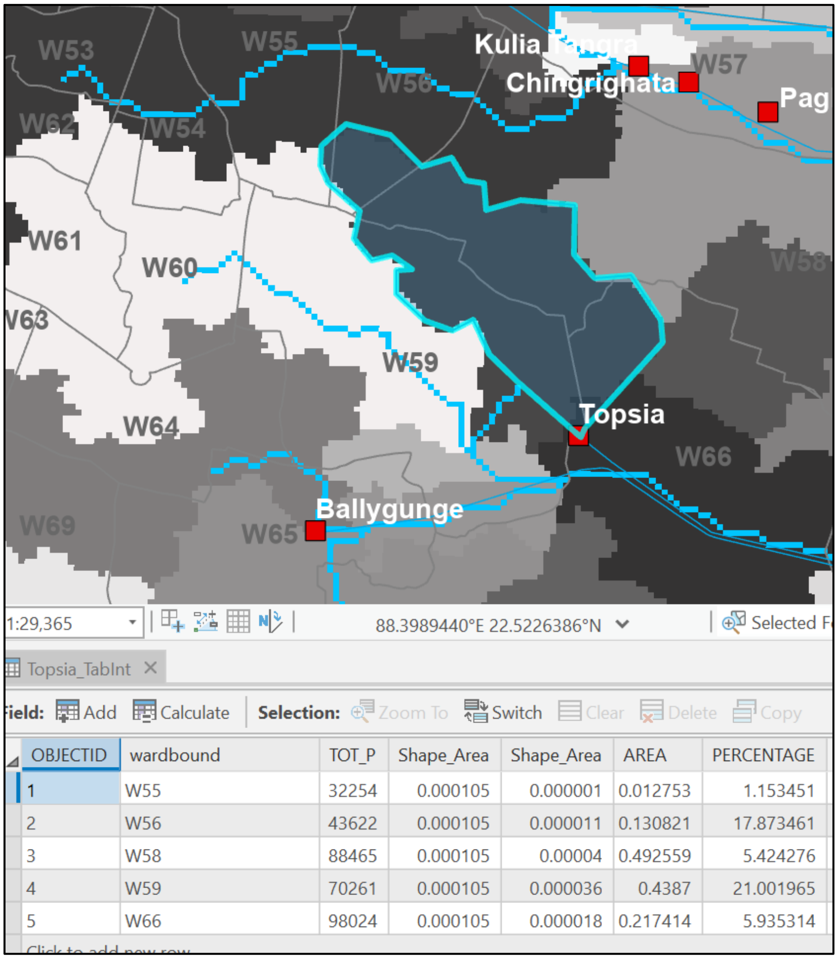
Step in the Process Flow Diagram:

> Using the direction of drainage indicated by the catchment areas, select wards that drain towards the pumping station and sum corresponding populations.

According to the Irrigation and Waterways Department's list of pumping stations with installed capacity figures, there is only one facility referred to as "Topsia" with a capacity of 65 cusecs[43]; however, referring to the official website of KMC, the list of pumping stations shows two pumping stations (Topsia DPS and Topsia Point 'A') with the same number of pumps (N = 7)[65]. After referring to the results from ES activities, it appears that only "Topsia (Old) DPS" was sampled and that Topsia Point 'A' was not sampled. This creates confusion because according to the literature, sewage is conveyed through high-level sewers to Topsia Point 'A' from Ballygunge and Palmer's Bridge stations[45]. Although it is uncertain form the literature whether these two stations are separate or one in the same, it is assumed that they are not due to the unique pumping station names on the KMC website. For this reason, the catchment delineation grid was used to guide population size calculations, and it was assumed that the sampled pumping station does not receive wastewater from Ballygunge PS or Palmer's Bridge PS. This assumption should be verified with local experts.

A zone was drawn to capture the upstream catchment area of the Topsia PS (Figure B.18). Although the catchment grid delineation map indicates a drainage pattern slightly southeast of this station, it was assumed that the direction of flow is generally towards the Topsia PS. The "Tabulate Intersection" tool was then used to determine the percentage of area that each ward shares with the upstream catchment area. These percentages were multiplied by the total population in each corresponding ward and summed to calculate the total population estimate. Therefore, the population size contributing to the Topsia (Old)

PS sample location is estimated to be 33,542 people; however, it is uncertain how much of this estimate overlaps with Topsia Point ‘A’ PS.



OBJECTID	wardbound	TOT_P	Shape_Area	Shape_Area	AREA	PERCENTAGE
1	W55	32254	0.000105	0.000001	0.012753	1.153451
2	W56	43622	0.000105	0.000011	0.130821	17.873461
3	W58	88465	0.000105	0.00004	0.492559	5.424276
4	W59	70261	0.000105	0.000036	0.4387	21.001965
5	W66	98024	0.000105	0.000018	0.217414	5.935314

Figure B.18 Upstream catchment area of the Topsia (Old) PS