
Haokai Zhao

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2024
Abstract


Haokai Zhao

Cities are home to more than half of the world’s population, and this figure is set to continue to rise amidst ongoing global urbanization trends. Against this backdrop, urban development is increasingly confronted with multifaceted challenges. These range from public health emergencies, exemplified by the COVID-19 global pandemic, to the environmental hazards driven by climate change, including extreme heat waves and more frequent severe storms. Confronted with these substantial risks, the urgency of devising and implementing strategies for sustainable and resilient urban development has become paramount. Given this context, the work presented in this thesis aims to advance understanding of some critical urban sustainability challenges, and to develop models, tools, and sensing systems that can support progress towards a more sustainable and resilient urban future.

The first part of the thesis focuses on the role and usage of urban parks during a global public health emergency. Urban parks became critical for maintaining the well-being of urban residents during the COVID-19 global pandemic. To examine the impact of COVID-19 on urban park usage, New York City (NYC) was selected as a case study, and SafeGraph mobility data, which was collected from a large sample of mobile phone users, was used to assess the change in park visits and travel distance to a park based on park type, the income level of the visitor’s census block group (visitor CBG) and that of the park census block group (park CBG). All
analyses were adjusted for the impact of temperature on park visitation, and the research work was focused primarily on park visits made by NYC residents. Overall, for the eight most popular park types in NYC, namely – Community Park, Flagship Park, Jointly Operated Playground, Nature Area, Neighborhood Park, Playground, Recreation Field/Courts and Triangle/Plaza – visits dropped by 49.2% from 2019 to 2020. The peak reduction in visits occurred in April 2020. Visits to all park types, excluding Nature Areas, decreased from March to December 2020 as compared to 2019. Parks located in higher-income CBGs tended to have lower reductions in visits, with this pattern being primarily driven by visits to large parks, including Flagship Parks, Community Parks and Nature Areas. All types of parks saw significant decreases in distance traveled to visit the park, with the exception of the Jointly Operated Playground, Playground, and Nature Area park types. Visitors originating from lower-income CBGs traveled shorter distances to parks and had less reduction in travel distances compared to those from higher-income CBGs. Furthermore, both before and during the pandemic, people tended to travel a greater distance to parks located in high-income CBGs compared to those in low-income CBGs. Finally, multiple types of parks proved crucial destinations for NYC residents during the pandemic. These included Nature Areas to which the visits remained stable, along with Recreation Field/Courts which had relatively small decreases in visits especially for lower-income communities. Results from this particular research study can support future park planning by shedding light on the different users of certain park types before and during a global crisis, where access to green spaces can help alleviate the human well-being consequences associated with mitigating the crisis, including the type of “lockdown” or limited mobility policies implemented in 2020 during the COVID-19 global pandemic.
The second part of the thesis investigates the role of urban greening and other land surface features in influencing the urban heat island effect in NYC. The urban heat island (UHI) effect describes the phenomenon whereby cities are generally warmer than surrounding rural areas. UHI effects can exacerbate extreme heat events, leading to an increase in heat-related illness and mortality. Here, the runoff coefficient was used as a numerical surrogate for urban greening, with lower runoff coefficients being associated with higher fractions of urban greening.

Using a high-resolution landcover GIS dataset developed for New York City (NYC), which classified the city into more than 13 million land patches, the runoff coefficient of land use across the entire city was mapped down to a resolution of 30m×30m, along with five other variables including surface albedo, distance to water bodies, land surface elevation, building density and building height. Daytime land surface temperature (LST) in summer was used as a surrogate for the UHI effect in NYC, and the work investigated the relationship between the runoff coefficient and LST. The work also examined the relationship between LST and the variables of surface albedo, distance to a water body, land surface elevation, building density and building height. Results indicate that runoff coefficient can explain a large portion of variability related to urban LST, with lower runoff coefficients (more greenery) being associated with lower LST. Use of the five other variables improves the predictability of LST, although the influence each variable has on LST varies with urban setting and context. The research work presented in this part of the thesis also shows the disproportionately higher exposure to urban heat in lower-income communities in NYC. The findings can be used to develop strategies to mitigate UHI effects in NYC and other cities around the world.

In the third part of the thesis, a wireless environmental sensing system is developed for monitoring urban green spaces, with demonstrated application for stormwater management. The
monitoring of urban green spaces, including monitoring of soil conditions and soil health, is crucial for sustainable urban development and ecological resilience. Leveraging advances in wireless environmental sensing, a LoRaWAN-based system capable of measuring air temperature/humidity, soil temperature and moisture, and soil moisture dynamics is designed and deployed across seven diverse urban green spaces for a full year at Columbia University’s Morningside Campus in New York City. The data collected by this sensing network reveals notable variations in soil moisture across the seven monitored sites, which are influenced by a combination of vegetation type, soil conditions, and physical settings. Monitored lawns consistently showed higher soil moisture levels due to their slower draining soil type, underlying concrete structures, and lower canopy rainfall interception and transpiration loss, whereas one monitored tree pit site with a more rapidly draining soil type showed significantly lower soil moisture throughout the study period, despite having comparable physical settings with another monitored site. Seasonal trends indicated lower summer moisture in some monitored areas due to increased evaporation and transpiration under high temperatures, while other areas maintained higher soil moisture as a result of frequent irrigations. Models were developed to quantify soil moisture response to rainfall events. It was found that the increase in soil moisture at each monitored site was highly dependent on the rainfall depth and the initial soil moisture. Overall, the results show that a range of diverse green spaces can help retain and drain storms up to certain sizes of 30-50mm. However, proactively designed soil drainage systems are needed to handle extreme storm events above 50mm. The study highlights the effectiveness of LoRaWAN technology in urban environmental monitoring and provides valuable insights into how different urban green spaces can contribute to stormwater management. The findings presented in this portion of the thesis demonstrate the instrumental role that monitoring, data analysis and
modeling can play in helping city planners and environmental managers optimize urban green spaces for ecological benefits and enhance urban resilience, including in the face of stressors such as climate change.

Overall, with its data-driven, evidence-based insights, this work contributes to the understanding of the multifaceted urban sustainability challenges in a changing environment, including public health emergencies such as the COVID-19 global pandemic, and climate change induced environmental hazards such as extreme heat events and more frequent severe storms. Alongside deepening understanding, the developed quantitative models and sensing technologies presented in this thesis offer practical solutions to support urban development towards a more sustainable and resilient future.
Table of Contents

List of Figures ......................................................................................................................... iii
List of Tables ........................................................................................................................... vi
Acknowledgments .................................................................................................................... vii

Chapter 1 – Introduction .........................................................................................................1

1.1 Motivation .........................................................................................................................1
1.2 Research Objectives .........................................................................................................8
1.3 Thesis Organization ............................................................................................................8

Chapter 2 – Change of urban park usage as a response to the COVID-19 global pandemic .........11

2.1 Introduction .......................................................................................................................11
2.2 Data and Methods ............................................................................................................14
2.3 Results ...............................................................................................................................17
   2.3.1 The types of parks chosen for analysis ................................................................. 17
   2.3.2 Park visits and visitors change rate ..................................................................... 18
   2.3.3 Travel distance to the parks ............................................................................. 22
2.4 Discussion .........................................................................................................................27

Chapter 3 – Quantifying the role of urban greening and other key factors in influencing the urban heat island effect ...........................................................................................................34

3.1 Introduction .......................................................................................................................34
3.2 Data and Methods ............................................................................................................37
3.3 Results and Discussion .....................................................................................................44
   3.3.1 Relationship between LST and runoff coefficient and other urban features ... 44
   3.3.2 Effect of surrounding runoff coefficients on LST ............................................ 47
   3.3.3 Effect of different types of water bodies on LST ............................................. 49
   3.3.4 Effect of land surface elevation on LST ......................................................... 50
   3.3.5 Effects of building features on LST ................................................................. 52
   3.3.6 The unequal exposure to the heat across income groups ............................... 55
3.4 Conclusions .......................................................................................................................56

Chapter 4 – A LoRaWAN-based environmental sensing network for urban green space monitoring with demonstrated application for stormwater management ........................................59

4.1 Introduction .......................................................................................................................59
4.2 System Design and Methods ..........................................................................................62
   4.2.1 Sensing system design ...................................................................................... 62
   4.2.2 Soil moisture sensor calibration ...................................................................... 64
List of Figures

Figure 2-1 Locations of parks in New York City. Each point represents an individual park. The number of parks and the median park area are shown for each park type........................................ 17

Figure 2-2 Park visits and visitors change rate by borough, which is calculated as the percent change of total monthly park visits/visitors in 2020 compared to 2019. The letters on the right of each figure are the Tukey HSD multi-group comparison results, the same letters indicate the boroughs belong to the same group................................................................. 19

Figure 2-3 Park visits and visitors change rate by park type, which is calculated as the percent change of total monthly park visits/visitors in 2020 compared to 2019. The letters on the right of each figure are the Tukey HSD multi-group comparison results, any common letter shared by two park types indicates that the two park types were found to belong to the same group................................................................. 20

Figure 2-4 NYC park visitors change rate by park type and by income level of park CBGs. The letters before the income groups are the Tukey HSD multi-group comparison results, any common letter shared by two income level groups indicates that the two groups were found to belong to the same group......................................................................................... 22

Figure 2-5 (a): Mean travel distance by park type. (b): Percentage change of mean travel distance by park type, with 95% CI error bars........................................................................................................... 24

Figure 2-6 (a) and (b): Mean travel distance by park type and by income level of visitor CBGs; The letters to the right of the mean travel distances are the Tukey HSD multi-group comparison results between income groups for each park type, any common letter shared by two income groups indicates that the two groups were found to belong to the same group. (c): Percentage change of mean travel distance by park type and by income level of visitor CBGs, with 95% CI error bars. ...................................................................................................... 25

Figure 2-7 (a) and (b): Mean travel distance by park type and by income level of park CBGs; The letters to the right of the mean travel distances are the Tukey HSD multi-group comparison results between income groups for each park type, any common letter shared by two income groups indicates that the two groups were found to belong to the same group. (c): Percentage change of mean travel distance by park type and by income level of park CBGs, with 95% CI error bars. ...................................................................................................... 26

Figure 3-1 Maps of runoff coefficient and other land surface features in New York City............. 43

Figure 3-2 LST vs runoff coefficient, with the OLS linear regression result................................. 45

Figure 3-3 XGBoost regression results: (a) testing accuracy, (b) feature impact on model output, and (c) feature importance by mean absolute SHAP value ......................................................... 46
Figure 3-4 CBG runoff coefficient of pixels in different temperature groups for (a) low runoff coefficient pixels, (b) medium runoff coefficient pixels and (c) high runoff coefficient pixels; means and medians were shown as triangles and orange lines. ........................................ 47

Figure 3-5 LST of pixels within 30m of different types of water bodies; means and medians were shown as triangles and orange lines. ................................................................. 49

Figure 3-6 LST of pixels in different elevation groups that were at least 90m away from water; means and medians were shown as triangles and orange lines. ............................... 50

Figure 3-7 CBG LSTs in different building density groups; means and medians were shown as triangles and orange lines. ......................................................... 52

Figure 3-8 LST of the building roof pixels in different albedo groups; means and medians were shown as triangles and orange lines. ......................................................... 53

Figure 3-9 (a) Very high building density CBGs colored by building height group; (b) LST of pixels in different building height groups for CBGs with very high building density; means and medians were shown as triangles and orange lines. ............................... 54

Figure 3-10 LST of pixels in different income groups; means and medians were shown as triangles and orange lines. ................................................................. 55

Figure 4-1 (a) System diagram and illustration of the four different vegetation settings used in study, (b) 3d-printed sensor box and system assembly, (c) LoRaWAN gateway, (d) weather station, (e) system in operation in the field (node 7), (f) real-time data monitoring dashboard. ............................................... 63

Figure 4-2 (a) Collected soil samples, (b) soil moisture sensor calibration curves, (c) soil texture estimation by sedimentation method (node 4). ......................................................... 66

Figure 4-3 Device locations and field photos showing sensor node settings. ................. 67

Figure 4-4 Distribution of soil moisture at each measurement site throughout the monitoring period of May 7th 2022 to May 7th 2023. ......................................................... 70

Figure 4-5 Distribution of soil moisture at each measurement site in different seasons. ....... 71

Figure 4-6 Increase in soil DoS as a response to rainfall amount; the star marks after the regression coefficients and the Spearman’s rank correlation coefficients indicate the significance level (***: p<0.001, **: p<0.01, *: p<0.05). ........................................ 73

Figure 4-7 Increase in soil DoS as a response to rainfall amount and initial soil DoS; the star marks after the regression coefficients indicate the significance level (***: p<0.001, **: p<0.01, *: p<0.05). ........................................................................................................ 75

Figure A-1 Models of park visits vs temperature ......................................................... 106

Figure A-2 (a) Income level of park CBGs by park type; (b) park area by park type. The letters inside the boxes are the Tukey HSD multi-group comparison results for each topic, any
common letter shared by two park types indicates that the two park types were found to belong to the same group. ................................................................. 111

Figure A-3 Timeline of the closure and reopening of NYC parks ................................. 113
Figure A-4 Timeline of the closure and reopening of NYC schools ............................... 113
Figure A-5 Example figures of raw visits data vs. normalized visits data by total visitors. (a) Flagship Park (b) Nature Area ................................................................. 114
Figure B-1 Map of the LoRaWAN signal survey in Upper Manhattan ......................... 115
List of Tables

Table 3-1 The major New York City land use categories (non-water type) and the associated runoff coefficients .................................................................................................................................................. 38
Table 3-2 Selected feature subsets for predicting LST based on SFS and BFS feature selection methods .................................................................................................................................................. 46
Table 4-1 Major components of the sensing system ........................................................................................................................................................................................................... 62
Table 4-2 Sites description and soil texture estimations for each site ........................................................................................................................................................................................................... 68
Table 4-3 Characteristic runoff coefficient by rain group ................................................................................................................................................................................................. 76
Table A-1 Summary of visit and visitor count by park type ........................................................................................................................................................................................................... 104
Table A-2 Relationship between the number of local NYC park visitors and temperature ......................................................................................................................................................................... 105
Table A-3 Paired t-test results for monthly temperatures in 2020 and 2019 ........................................................................................................................................................................................................... 105
Table A-4 Summary of the mean and change of travel distance by park type ........................................................................................................................................................................................................... 108
Table A-5 Summary of the mean and change of travel distance by park type and by income level of visitor CBGs ........................................................................................................................................................................................................... 109
Table A-6 Summary of the mean and change of travel distance by park type and by income level of park CBGs ........................................................................................................................................................................................................... 110
Table A-7 Park classification standard by the NYC Department of Parks and Recreation (DPR) ........................................................................................................................................................................................................... 112
Table A-8 Pearson’s correlation coefficient between raw visits and normalized visits by total visitors for each park type ........................................................................................................................................................................................................... 114
Acknowledgments

From the impacts of the COVID-19 global pandemic to the frontiers of scientific discovery and engineering advancement, it has been a challenging yet rewarding journey. While personal endeavor and perseverance played a pivotal role in achieving this milestone, as I culminate this journey that is my doctoral dissertation, I would like to express my deep gratitude to the individuals who have supported, inspired, and guided me along this path.

First, I extend my sincere thanks to my advisor Trish Dr. Patricia Culligan. The pandemic and the remote meetings have posed many difficulties in our efforts to lead to my PhD, however, your solid expertise, patient mentorship, and unwavering support have been the pillars of my academic growth. I’m not sure whether I am arguably among your students who argues with you the most, while I’ll definitely miss those constructive, insightful and fruitful discussions that we had, which has not only helped shape this research, but has also profoundly inspired my thinking as a researcher. I would also like to thank Dr. Brian Mailloux, John Dr. Ioannis Kymissis, Dr. George Deodatis and Dr. Markus Schläpfer for serving on my defense committee, your feedback and guidance during this crucial examination of my work were very valuable, it was my honor to have you as part of this important milestone in my academic journey.

I am also very grateful to my collaborators, including Kevin Kam and John Dr. Ioannis Kymissis from the Columbia Laboratory for Unconventional Electronics (CLUE), Dr. Brian Mailloux and Dr. Elizabeth Cook from the Environmental Science Department of Barnard College, and Jan Janak from the Columbia University Internet Real-Time (IRT) Lab. Your insights, perspectives, and contributions have been integral to the realization of this work. Collaborating with you has been a pleasant and learning experience that I will always cherish.
To my parents Yuxin Zhao and Huihui Wang: the pandemic and other barriers during my PhD study separated us for long; while there was always a difference of opinion between two generations as we grew up in different environments and stages of social development, when facing all these years’ challenges as a family, I was really pleased of our deepened mutual understanding and comprehended more about family as a natural emotional bond. Thank you, for your love and support; although I know you care more about me than this degree, this could not come true without you.

To my friends including Zhixiang Zhang, Shasha Zhu, Hanqing Fan, Aijia Hao, Kevin Kam, Diandian Zhao, Jinming Zhang, Weiwei Zhan, Mengyao Shen, Gurpreet Singh, Barhador Bahmani, Lingting Shi, Xin Wang, MinJung Kim, Xuechun Bai, Bowen Zhang, Shuqi Zhang, Jingbin Cao, Zhaoli Wen, Yu Huang, Caroline Jiang, Haoqi Zhao, Caroline Yu, Ziniu Wang, Xinyi Lyu, Yierfan Maierdan, Huyi Zhang, Kai Zhou, Yuxuan Huang, and my soccer teammates including Yinuo Jin, Yuxiang Liu, Zicong Huang, Yunran Zhou, Yunsai Zhang, Peng Lin and Zhenhao He: just by typing your names made a smile on my face, haha. From sharing knowledges in all of your interesting fields, playing sports and music that kept us energized, to hanging out in the city and exploring the wild, or just doing marathon talks on the phone to share stories in our lives; thank you for believing in me and inspiring me through all the ups and downs. I felt really fortunate to know you and have you in my life, and I look forward to many more years of friendship to come.

A special thank you to Ricky Gonowrie, Timothy Doherty and the Columbia Facilities team for always being kind and supportive to help me set up my sensing network. I would also like to thank the CEEM department and the Carleton Lab staff including Scott Kelly, Nathalie
Benitez, Michael Smith, Dr. Liming Li, Freddie Wheeler Jr. and Dr. William Hunnicutt, your professionalism, kindness and support have always made our department and Columbia feel like home.

To my Latino/Latina friends in my apartment building and my office building, including Joan, Elvin and Jilda, and those in the Strokos Deli – gracias por sus saludos diarios y por ayudarme a practicar mi español. Sus amables palabras y sonrisas han sido una fuente de alegría diaria.

Lastly, echoing the theme of this dissertation, I hope that cities can exhibit the same resilience to sustainability challenges as PhD students show on their doctoral journeys. Just as we adapt, learn, and grow through our experiences, I hope we can make our urban environments continue to evolve and thrive in the face of challenges. Cheers.
It is to see the world as it is, and to be brave.
1.1 Motivation

Global urbanization has been taken place at a dramatic pace in recent decades. The proportion of the world’s population that live in urban areas has grown from 30% in 1950 to 55% in 2018, and this number is projected to be 68% by 2050. As a multifaceted process that involves population migration, landcover change, urban agglomeration, socio-economic and other factors, urbanization has largely promoted fiscal, cultural and societal development and has also been recognized as an important way to improve human well-being. However, extensive human activities carried out during the urbanization process have also brought a set of adverse effects to the urban environment, such as air and water pollution, urban flooding, urban heat island effects, excessive anthropogenic greenhouse gas emissions, and so on. Furthermore, the current rapidly changing environment poses a series of challenges to sustainable urban development, including management of public health emergencies, such as the COVID-19 global pandemic, and management of the environmental hazards induced by climate change, such as extreme heat waves and more frequent severe storms.

The work presented in this thesis is motivated by the goal of advancing knowledge that can support progress towards sustainable urbanization, with an emphasis on the role that urban greening and green spaces can play in fostering better urban futures. While the role of urban green spaces has been recognized in promoting the health and well-being of city dwellers, in helping to mitigate urban heat island effects, and in reducing stormwater runoff and flooding during of wet-weather flow, gaps in knowledge still exist with respect to the type and
location of green spaces needed for benefit optimization. This work sought to fill these gaps by examining: 1) urban park usage during the COVID-19 pandemic, as a means of understanding how park type and location supported the physical and mental health of urban residents during lockdown restrictions, 2) quantifying how urban greening, as described by the runoff coefficient of an urban area, can contribute to urban cooling, and 3) designing and implementing a Long Range Wide Area Network (LoRaWAN) technology based monitoring network on an urban campus to demonstrate its effectiveness in long-term urban soil health monitoring and in elucidating the relationship between urban green space conditions and stormwater runoff for a range of storm events. The following paragraphs describe the motivation and work conducted under each of these three areas in more detail.

The COVID-19 pandemic has had a profound impact on urban life, reshaping the daily realities of city dwellers across the globe. The imposition of lockdowns and social distancing measures fundamentally altered urban dynamics, restricting movement and limiting access to public spaces and amenities. This led to a significant shift in how urban residents interacted with their environments, emphasizing the importance of local green spaces for mental and physical well-being. The pandemic also accelerated the adoption of remote work, reducing commuting and changing the usage patterns of public transport and city centers. Economically, it intensified inequalities, as lower-income communities and certain sectors faced disproportionate challenges. Socially, it fostered a renewed sense of community in some areas, while exacerbating feelings of isolation in others. In essence, the pandemic has not only disrupted urban life but also prompted a reevaluation of what it means to live in and design for resilient, sustainable urban communities.
Urban parks, recognized for their diverse ecosystem services and significant contributions to the physical and mental well-being of urban residents\textsuperscript{16,17,37,38}, gained heightened importance in the context of the COVID-19 pandemic. The pandemic's constraints, notably the dramatic curtailment of recreational opportunities and heightened concerns over personal and public health, transformed these green spaces into crucial destinations for city dwellers. Amid these challenges, researchers have endeavored to comprehend the role of parks during the pandemic using various data and methodologies\textsuperscript{24,26,39–41}. However, prior studies in this area often employed broad categorizations for parks, such as 'urban' or 'nature' parks, and faced limitations regarding population sample size, potentially compromising the representativeness of their findings.

SafeGraph, a location-based analytics company established in 2016, provides a solution to the above challenges. By collecting and compiling anonymized GPS data from mobile phone applications, SafeGraph provides extensive insights into population mobility patterns at a fine spatial scale\textsuperscript{42,43}. This rich dataset has been instrumental in multiple studies exploring human mobility, spatial interactions, and their correlations with the spread of COVID-19\textsuperscript{44–46}. Utilizing SafeGraph's comprehensive data, alongside other pertinent datasets, the research presented in this thesis sought to elucidate the changes in visitation patterns to different types of urban parks, using New York City as a case study, and to examine the variations in these changes by socio-economic factors. The goal of the work was to derive insights that can be used to enhance the resilience and functionality of urban parks as integral components of urban infrastructure systems, and most especially during future public health emergencies.
The Urban Heat Island (UHI) effect, a phenomenon where urban areas experience significantly higher temperatures than their rural surroundings, is primarily attributed to the extensive use of heat-absorbing materials in urban construction, such as concrete and asphalt, and the relative lack of vegetation. The exacerbation of UHI is further compounded by climate change, leading to more frequent and severe heat events in urban settings. These heightened temperatures can have a multitude of impacts on urban life and the environment.

First, extreme heat events pose significant health risks, especially to vulnerable populations such as the elderly, children, and those with pre-existing health conditions. They are associated with an increase in heat-related illnesses and mortalities. Secondly, UHI intensifies air pollution and smog formation, exacerbating respiratory problems among urban residents. Thirdly, the increase in temperatures contributes to higher energy demands, particularly for cooling, which can strain electrical grids and increase greenhouse gas emissions. Socioeconomically, lower-income communities often face greater exposure to UHI effects due to inadequate access to green spaces and cooling facilities. This disparity highlights the need for equitable urban planning that prioritizes sustainable designs to mitigate the UHI effect and adapt to the ongoing challenges posed by climate change.

Urban green spaces have been highlighted in the mitigation of the UHI effect, often creating an ‘Urban Cooling Island’ (UCI) effect in cities. Many relevant studies have used remote sensing data to explore how greenery cools urban environments by categorizing the urban landscape into various land use/land cover (LULC) types and analyzing the impact of green space composition and configuration on land surface temperature (LST). However, the complexity of urban landscapes, where a single grid cell or land patch often contains multiple
LULC types, has posed challenges in quantifying factors that influence the LULC-LST relationship when LULC is treated as a single categorical variable.

To address this challenge, the work presented in this thesis examines how multiple factors influence UHI, introducing an approach using the runoff coefficient as a quantitative indicator of greening level in an urban area. The runoff coefficient, which measures the proportion of rainfall becoming surface runoff on a scale of 0 to 1, varies distinctly across different urban areas, and provides a nuanced reflection of urban greening levels. Specifically, urban areas with high runoff coefficients represent hardscape, while areas with low runoff coefficients represent green and other spaces that are able to effectively absorb rainfall. Focusing on New York City (NYC), a densely populated area with diverse land cover and socio-economic demographics, daytime LST in summer is used to represent the UHI effect. The influence of the runoff coefficient on LST, along with five other land surface features including surface albedo, distance to water bodies, land surface elevation, building density, and height, is quantified and used to develop models to predict LST at a scale of 30m×30m which can inform heat mitigation strategies. Crucially, the study also demonstrates the disparities in heat exposure among different income groups in NYC, highlighting the importance of heat mitigation strategies that address the needs of socially vulnerable communities.

Monitoring urban green spaces, particularly focusing on soil health and conditions, is vital for maintaining the health and functionality of these essential urban assets\textsuperscript{75,76}. Healthy soils are pivotal in supporting robust vegetation, which enhances the ecological and aesthetic value of urban areas\textsuperscript{77–79}. They play a key role in stormwater management, absorbing and filtering rainwater, thereby mitigating flood risks and improving water quality\textsuperscript{80–82}. This natural sponge-
like property of healthy soils in green spaces is increasingly important as cities confront more frequent and severe storms due to climate change\textsuperscript{83}. Effective management and monitoring of soil conditions in these spaces, as well as an in-depth understanding of soil moisture dynamics in response to rainfall events, are critical to ensuring that they maintain their capacity to manage stormwater and therefore properly contribute to urban resilience and sustainability.

Traditional methods for soil condition measurements, such as laboratory tests and in-situ measurements with various instruments\textsuperscript{84,85}, while reliable, often come with complexities in installation and high maintenance costs, limiting their practicality for widespread use in diverse urban settings. Recent advancements in environmental sensing technologies have opened up new opportunities for monitoring soil conditions in urban environments\textsuperscript{86,87}. Long Range Wide Area Network (LoRaWAN) technology, known for its low power consumption, long communication range, and ease of deployment, has become increasingly popular\textsuperscript{88,89}. LoRaWAN-based systems have found applications in various environmental monitoring areas, including air quality\textsuperscript{90,91}, weather conditions\textsuperscript{92}, tree health\textsuperscript{93}, and soil conditions\textsuperscript{94}. However, most existing studies utilizing this technology have focused on short-term data collection and primarily presented raw observational data, lacking in-depth analysis or model development. Despite its potential, a comprehensive LoRaWAN-based soil sensing network, specifically designed and tested for diverse urban environments, remains to be established.

In response to these challenges, the goal of this area of work was to develop a LoRaWAN-based environmental sensing system capable of measuring a series of environmental parameters, with a particular focus on soil moisture dynamics. Additionally, the plan was to enhance the system's reliability through laboratory calibration and deploy it across different
urban green spaces, showcasing its adaptability to varied urban contexts. The system was also to be operational over a full year, allowing for comprehensive data collection. To demonstrate the utility of the system, initial data analysis was to focus on understanding soil moisture trends and quantitatively modeling soil moisture response to rainfall events, as well as relationships between rainfall and runoff under different green space conditions. This work not only demonstrated the practical application of LoRaWAN technology in a real-world urban setting, but also deepened understanding of soil moisture and stormwater runoff dynamics in urban green spaces. Eventually, research conducted to support this part of the thesis aimed to provide crucial insights into monitoring strategies that can support effective green space management, with the ultimate goal of allowing cities to optimize the ecological benefits of urban green spaces and to enhance urban resilience against challenges like climate change.
1.2 Research Objectives

The primary objective of this dissertation is to advance the understanding of three critical urban sustainability challenges and to create insights and tools for addressing them. Specifically, this thesis aims to:

1) Investigate the changes in urban park visitation patterns during the COVID-19 pandemic, and disaggregate the changes by park location, park type and socio-economic factors;

2) Quantify the impact of urban greening and other land surface features on the Urban Heat Island effect, employing the runoff coefficient as an urban greening indicator;

3) Develop and implement a LoRaWAN-based wireless environmental sensing system for long-term soil conditions monitoring; use the system to examine soil moisture dynamics and quantify its response to rainfall events.

1.3 Thesis Organization

This thesis addresses the three research objectives in three consecutive chapters, from 1) the importance of urban parks during the COVID-19 pandemic, 2) the role of urban greening in mitigating excessive urban heat, to 3) intelligent environmental sensing in supporting stormwater management. In a separate concluding chapter, the thesis summarizes the overall findings and outlines avenues for future research. A brief description of the chapters following this introductory chapter is provided below.
Chapter 2 investigates the impact of the COVID-19 pandemic on urban park usage in New York City during 2020, utilizing SafeGraph mobility data derived from mobile phone users, as well as other geospatial and socio-economic datasets. The study assessed changes in park visitation and travel distances based on park type, and the income levels of both visitors and parks' locations. The findings reveal an overall significant decrease in park visits during 2020, compared to 2019, with variations across different park types and socioeconomic demographics. The study highlights how urban parks served as vital spaces for residents during the pandemic, with specific parks being identified as important destinations, offering insights for future park planning in crisis scenarios.

Chapter 3 examines the influence of urban greening and other land surface features on the urban thermal environment, specifically focusing on the Urban Heat Island (UHI) effect in New York City. The study uses the runoff coefficient as an indicator of urban greening, with lower coefficients corresponding to higher levels of greenery. Utilizing a high-resolution LULC GIS dataset of NYC, the research maps the runoff coefficient and its relationship with daytime LST in summer, a proxy for UHI, at a scale of 30m×30m. The study also incorporates other variables such as surface albedo, distance to water bodies, land surface elevation, and built environment characteristics to understand their impact on LST. Additionally, the exposure to urban heat across different income groups was assessed. The findings establish both qualitative and quantitative relationship between urban greening and other land surface features vs LST, providing insights for developing targeted UHI mitigation strategies.

Chapter 4 details the development and deployment of a wireless environmental sensing system for monitoring urban green spaces, emphasizing its application in stormwater
management. Utilizing a LoRaWAN-based system, the study measured a series environmental parameters and soil moisture dynamics across seven diverse urban green spaces at Columbia University’s Morningside Campus in New York City over a full year. The collected data revealed significant variations in soil moisture influenced by vegetation types, soil conditions, and physical settings. The study quantified the responses of soil moisture to rainfall events, highlighting how diverse green spaces can manage stormwater effectively up to certain rainfall sizes. However, it also revealed the need for implementing more proactive green space drainage systems for the management of extreme rainfall events. The study underscores the importance of monitoring, data analysis, and modeling in optimizing urban green spaces for ecological benefits and enhancing urban resilience, particularly against climate change stressors.

The concluding chapter of the dissertation, Chapter 5, presents the key findings and contributions, and proposes future research directions. References and appendices are provided at the end of the dissertation.
2.1 Introduction

Urban parks provide a variety of ecosystem services\textsuperscript{37,38,95}, as well as physical and mental health benefits to urban residents\textsuperscript{16,17,96}. The COVID-19 global pandemic drastically altered people’s mobility patterns\textsuperscript{22,23,97}, especially during the first several months when restrictions were implemented by governments to combat the spread of the disease. Such restrictions included stay-at-home orders, the closure of non-essential businesses, the cancelation of public events and in-person schooling, social distancing, and travel restrictions, etc.\textsuperscript{24,98,99} With the many challenges imposed by the pandemic, which included dramatically reduced recreational opportunities as well as widespread concerns about personal and public health, urban parks, which were one of the few places that urban dwellers were allowed to visit outside their homes, became important destinations.

In order to understand the role of parks during the COVID-19 pandemic, a variety of data and methods have been used by researchers to conduct relevant studies: including carrying out field surveys\textsuperscript{24,39,100}, recruiting civic scientists to make observations\textsuperscript{40,101}, collecting geotagged data from social media\textsuperscript{25,41,102,103}, and acquiring data from recreational tracking apps\textsuperscript{26}. Decreased visits to urban greenspaces in central London was reported\textsuperscript{103}, which could be attributed with working from home restrictions; similarly, in a study conducted in multiple cities across North Carolina, 56% of survey respondents indicated that they had ceased or reduced their use of parks, with geo-tracked park visits dropping by 15%\textsuperscript{100}. Several studies reported increased visits to urban greenspaces and nature parks away from city centers\textsuperscript{24,26,103}, while one study conducted in
four Asian cities indicated people’s preference for large nature parks close to city centers. The heightened appreciation for nature and the raised awareness of its importance have been highlighted by many studies. Nonetheless, the parks studied were usually only loosely categorized, such as urban parks and nature parks; and one major concern about methods used in these studies, is that the data might not represent the general population very well because of the limited sample sizes.

SafeGraph, a location-based product company founded in 2016, has been collecting and compiling anonymous GPS data from mobile phone apps, and aggregating these data to provide information on people’s mobility patterns. SafeGraph’s primary product is a places dataset of millions of points of interests (POIs) in the United States and Canada, providing details like business name, address, category, and geographic coordinates for each POI. They also provide a patterns dataset with anonymized, aggregated foot traffic data for those POIs. Because of the pervasiveness of smart phones in modern life, such data provides an opportunity for analyzing the mobility patterns of the general public. Several studies have employed SafeGraph data to investigate patterns of human mobility and spatial interaction, as well as their relationships with the spread of COVID-19.

The goal of this study was to use the SafeGraph dataset to understand how park usage changed during the early phases of the COVID-19 pandemic, including by park location, detailed park type, and the socio-economic level of the park visitors and of the neighborhood of the park. Given the extensive size of the mobility dataset, we selected New York City (NYC) as our study area. NYC is representative of many dense, urban environments and was one of the major cities affected by the first waves of the pandemic. We examined changes in the number of visits to
parks, the number of individual visitors to parks and the change in travel distance to parks. Our focus was primarily on park usage by NYC residents, rather than visitors to the city. Understanding how park usage changed during the pandemic is considered critical to the planning and management of urban public spaces in post-COVID cites, and in an era where future pandemics, and other crises that might lead to public “lockdowns”, cannot be ruled out.
2.2 Data and Methods

The mobility data were obtained from SafeGraph\textsuperscript{42} for the time period of January 2018 to December 2020. SafeGraph aggregates anonymized location data from numerous applications in order to provide insights about points of interest (POIs) that people visit, via the SafeGraph Community. To enhance privacy, SafeGraph excludes census block group (CBG) information if fewer than two devices visited an establishment in a month from a given census block group. As examined in our analysis that comparing 2020 data to 2018 yielded similar results to 2019, we chose to present comparisons of data from 2020 to 2019 for clarity.

SafeGraph’s \textit{places} dataset contains a variety of information for each POI in its product suite, including the location name, address, latitude and longitude coordinates, category, brand, and additional details. To select those POIs of our interest, we obtained the NYC Open Space Parks Data\textsuperscript{107}, which is a vector GIS dataset that provides information on park characteristics such as location, boundary, and park type. We then used the ArcGIS Pro software to spatially select those POIs within park boundaries. The NYC Open Space Parks Dataset was chosen because of its comprehensive coverage of the parks in the city, and the NYC Parks Department classified the parks into distinct types based on park features, service area and size, etc. (Table A-7).

SafeGraph designations don’t always match with NYC Parks Department designations. Some studies have focused on the SafeGraph designated POIs with type “Nature Parks and Other Similar Institutions”, as this approach unified the selection of parks across multiple cities\textsuperscript{108}; however, they excluded some park types such as playgrounds and greenspaces that may be coded under other categories in the SafeGraph POIs dataset, such as “Museums, Historical Sites, and
Similar Institutions” and “Other Amusement and Recreation Industries”, etc. Our approach included all these POIs that fall in the government-designated park boundaries and enabled comparisons between different types of parks in our studied urban area.

The monthly patterns dataset from SafeGraph contains visitors’ mobility information and is organized by POI. As mentioned in the types of parks chosen for analysis section, we calculated four types of visits: 1) all visits: to calculate the number of visits to each park, we summed raw_visit_counts to all the POIs within a park; 2) all visitors: raw_visitor_counts to all the POIs within a park were summed; 3) US visitors: the visitor_home_cbgs column provides the number of visitors to the POI from each US CBG based on the visitor's home location, they were summed to determine US visitors; 4) NYC local visitors: the visitors who are from a CBG within NYC were summed as the number of NYC local visitors. When deriving travel distances, we obtained the US Census Block Group (CBG) Boundaries GIS data\textsuperscript{109} from the US Census Bureau, then we calculated the distance between the centroid of a visitor’s home CBG and the destination POI.

Temperature has been reported as a vital factor influencing park visitation\textsuperscript{110–112}. To adjust visits data by temperature, NYC daily and monthly climate data were downloaded from the NOAA climate database\textsuperscript{113} for the year of 2018 to 2020. The data includes weather information from multiple weather stations throughout the city, which were then averaged to represent the citywide mean temperature. The relationship between monthly park visits and monthly mean temperature was confirmed using Pearson’s correlation coefficient. Overall, the number of park visits was found to be highly correlated with temperature (Pearson's R = 0.78, p <0.001) for the time period from January 2018 to February 2020 (Table A-2), which was prior to
the implementation of the COVID-19 policies restricting people's mobility in NYC. Except for jointly operated playgrounds, the relationship holds for all park types.

To adjust park visitation by temperature, we first conducted a paired t-test to compare NYC daily mean temperature in each month between 2019 and 2020. The temperature in 5 months was found to be significantly different in the two years (Table A-3), indicating the necessity for correction. Later, we built several models using the least squares method and the Gaussian Process method to model the relationship between park visits and temperature. The best one, a third-degree polynomial model with an $R^2$ value of 0.72, was then used to adjust park visits data (Equations A-1 and A-2).

To investigate the impact of socioeconomic factors, the American Community Survey's annual average per capita income data at the CBG level for the year 2019 were obtained. We defined the visitor census block group (visitor CBG) as the visitor's home CBG, which was provided by the `visitor_home_cbgs` column in the `monthly patterns` dataset; and we defined the park census block group (park CBG) as the CBG in which the park was located or was the closest to, which was determined using ArcGIS Pro software by attributing the nearest CBG to each POI.

The income data were then combined with park visits data based on visitor CBGs and park CBGs, and the income groups were determined by the per capita income level of these CBGs, with the three terciles serving as the cutoff numbers for lower-, middle- and upper-income levels. For group comparisons, e.g. visits change rate between boroughs/park types, visits change rate between income groups for each park type, and travel distance between income groups, etc. Tukey’s HSD post-hoc test was used.
2.3 Results

2.3.1 The types of parks chosen for analysis

The total number of park visits determined from SafeGraph points of interest (POIs) located in NYC parks was 20,913,290 in 2019, but only 10,279,798 in 2020, representing a decrease of 49.2 percent.

![Locations of parks in New York City. Each point represents an individual park. The number of parks and the median park area are shown for each park type.](image)

There are 18 types of parks listed in the NYC Open Space Parks Data\textsuperscript{107}. However, the top eight types of parks accounted for 91.35 % of total park visits in 2019 and 92.17 % in 2020, respectively, and are thus the focus of this study (Figure 2-1). These parks are classified by the NYC Department of Parks and Recreation\textsuperscript{107} as 1) Community Park, 2) Flagship Park, 3) Jointly Operated Playground, 4) Nature Area, 5) Neighborhood Park, 6) Playground, 7) Recreation Field/Courts, and 8) Triangle/Plaza. The detailed classification standard can be found in Table A-7.
Four metrics associated with the number of park visits and the number of park visitors were calculated, namely 1) all visits: the total number of visits from all visitors; From the SafeGraph documentation, the duration of a visit must last at least 4 minutes, and there could be multiple visits from a single visitor during the time period when the data were collected; 2) all visitors: the total number of unique visitors, regardless of their origin; 3) US visitors: the total number of unique visitors whose home locations are within the US; 4) NYC local visitors: the total number of unique visitors whose home locations are within NYC. Since temperature has been reported as a vital factor influencing park visitation, we corrected the data for the effects of temperature, as described in the Data and Methods section. The total numbers of these four types of visit/visitor counts by park type, and after the temperature correction, are summarized in Table A-1.

2.3.2 Park visits and visitors change rate

2.3.2.1 Park visits and visitors change rate by borough

We examined the park visits and visitors change rate in each NYC borough by computing the total number of park visits or visitors in a month in that borough, then calculating the percentage change in 2020 visits (or visitors) compared to 2019 (i.e., (the visits in 2020 - the visits in 2019) / the visits in 2019). Manhattan was divided into lower Manhattan and upper Manhattan using 86th street as a divide. Results for total visits, total visitors, US visitors, and NYC visitors are shown in Figure 2-2.

Starting from March 2020, the parks in all boroughs experienced decreased total visits (Figure 2-2a). April 2020 was the month with the greatest percentage decrease in visits compared to 2019, then visits slowly increased as the months progressed. Lower Manhattan had the
greatest decrease in park visits from March to December (overall -61.1%, with a maximum of -86.6% in April 2020), while Staten Island experienced the smallest decrease (overall -20.3%, with a maximum of -57.5% in April 2020). All other boroughs experienced similar changes in visits, and shared a similar trend through time. All visitors (Figure 2-2b), U.S. visitors (Figure 2-2c), and NYC visitors (Figure 2-2d) had the same pattern with the greatest decrease in April, followed by a slow rebound with progressing time. Again, lower Manhattan had the largest decrease in unique visitors while Staten Island the smallest (Figure 2-2b-d).

2.3.2.2 Park visits and visitors change rate by park type

We examined the park visits and visitors change rate across the eight selected park types, by computing the total number of visits or unique visitors in a month to each park type and then calculating the percentage change of visits/visitors in 2020 compared to 2019 (Figure 2-3). There was a decrease in all types of visits and visitors to all eight park types across the city when
comparing 2019 to 2020 for the months of March to June (Figure 2-3a-d). For NYC local visitors, Triangle/Plazas (overall -62.9%, with a maximum of -82.9% in April 2020) and Flagship Parks (overall -57.0%, with a maximum of -78.7% in April 2020) had the largest decrease, followed by Jointly Operated Playground, Playground, Community Park, Neighborhood Park, then Recreation Field/Courts. Nature Areas had the smallest decrease in the number of NYC local visitors (overall -3.6%, with a maximum of -44.5% in April 2020) with some months even showing an increase. Beginning in June, the number of NYC local visitors to Nature Areas returned to about the same level as 2019 and even increased in some months (with a maximum increase of 29.0% in July 2020) (Figure 2-3d). The other three types of visits/visitors shared similar trends as NYC local visitors (Figure 2-3a-c).

![Figure 2-3](image)

Figure 2-3 Park visits and visitors change rate by park type, which is calculated as the percent change of total monthly park visits/visitors in 2020 compared to 2019. The letters on the right of each figure are the Tukey HSD multi-group comparison results, any common letter shared by two park types indicates that the two park types were found to belong to the same group.

In order to better understand the needs and park usage of local urban residents, we focused our remaining analyses on data for NYC residents (also subsequently referred to as NYC...
local visitors) only. We also defined the visitor census block group (visitor CBG) as the home census block group where a visitor lived; and defined the park census block group (park CBG) as the census block group that a park was in, or was the closest to.

2.3.2.3 Park visitors change rate by both park type and by income level of park CBGs

The CBGs (neighborhoods) that surround parks were divided into three income groups: lower, middle and upper, based on the per capita income. The results for park visits change rate between 2019 and 2020 for each of the analyzed park types are provided in Figure 2-4.

All eight park types saw decreased NYC local visitors regardless of the park CBG income level (Figure 2-4). Overall, parks in lower-income neighborhoods experienced statistically greater decreases in NYC local visitors than those in upper-income neighborhoods. No trend in visits change rate with income level was observed for Jointly Operated Playground, Neighborhood Park, Playground, and Triangle/Plaza. Community Parks and Nature Areas showed greater reductions in NYC local visitors in lower-income neighborhoods but showed no difference between middle- and upper-income neighborhoods. Flagship Parks showed greater reductions in NYC local visitors in lower- and middle-income neighborhoods. The outlier to the overall trend is Recreational Field/Courts, which showed greater reductions in NYC local visitors in upper-income neighborhoods than in lower-income neighborhoods.
Figure 2-4 NYC park visitors change rate by park type and by income level of park CBGs. The letters before the income groups are the Tukey HSD multi-group comparison results, any common letter shared by two income level groups indicates that the two groups were found to belong to the same group.

2.3.3 Travel distance to the parks

The travel distance of visitors was used to examine how the travel behavior of NYC residents to parks changed during the early stages of the COVID-19 pandemic. In this section,
the mean travel distances were computed for the time period from March to December in 2019 and 2020, as the major outbreak of the pandemic and the associated travel restrictions began in March 2020.

2.3.3.1 Change in travel distance by park type

Overall, the mean travel distance of NYC residents to all parks reduced from 5.9 km in 2019 to 5.1 km in 2020 over March to December, representing a change of -13.2% (95% CI: -13.4%, -13.1%). In 2020, there was a significant decrease in travel distance compared to 2019 for all study park types except for the Jointly Operated Playground, Playground, and Nature Area park types (Figure 2-5).

Before the pandemic, the mean travel distances to the Triangle/Plaza and Flagship Park types were the longest, both averaging 7.1 km – over March to December in 2019; while the travel distances to the Playground and Jointly Operated Playground were the shortest, averaging 5.0 km and 4.5 km, respectively. The Nature Area, Jointly Operated Playground and Playground park types experienced a smaller decrease than average or even a slight increase, which were -1.2% (95% CI: -2.0%, -0.3%), -1.9% (95% CI: -2.5%, -1.3%) and 1.1% (95% CI: 0.6%, 1.7%), respectively. All other types of parks experienced a greater reduction in travel distance (Table A-4).
2.3.3.2 Change in travel distance by both park type and by income level of visitor CBGs

The overall mean travel distances of NYC residents from lower-, middle- and upper-income level CBGs were 5.3 km, 6.5 km, and 6.0 km, respectively, from March to December in 2019; and were 4.7 km, 5.6 km, and 5.0 km, respectively, in the same period in 2020 (Figure 2-6a,b). In general, people from lower-income CBGs traveled a statistically shorter distance to parks than those from middle-income and upper-income CBGs in both 2019 and 2020. This pattern was common across all types of parks, except for Nature Areas and Triangle/Plazas, to which visitors from upper-income CBGs traveled the shortest distance.

Overall, visitors from higher income CBGs had the greatest reduction in travel distance (Figure 2-6c). The percentage change of travel distance for visitors from lower-income, middle-income and upper-income CBGs are -10.7% (95% CI: -11.0%, -10.4%), -13.9% (95% CI: -14.2%, -13.7%), and -15.8% (95% CI: -16.2%, -15.5%), respectively. (Table A-5)
The specific changes varied by park type. For Community Park, Flagship Park, Jointly Operated Playground, Nature Area and Triangle/Plaza park types, the visitors from upper income level CBGs experienced the greatest percentage reduction in travel distance. While for Recreation Field/Courts, the visitors from upper income level CBGs had the smallest percentage reduction in travel distance.

Figure 2-6 (a) and (b): Mean travel distance by park type and by income level of visitor CBGs; The letters to the right of the mean travel distances are the Tukey HSD multi-group comparison results between income groups for each park type, any common letter shared by two income groups indicates that the two groups were found to belong to the same group. (c): Percentage change of mean travel distance by park type and by income level of visitor CBGs, with 95% CI error bars.

2.3.3.3 Change in travel distance by both park type and by income level of park CBGs

The mean travel distances to parks located in lower-, middle- and upper-income level CBGs were 5.7 km, 5.6 km, and 6.3 km, respectively, from March to December in 2019, and were 4.9 km, 4.7 km, and 5.6 km, respectively, during the pandemic from March to December in 2020 (Figure 2-7a,b). In general, people tended to travel a statistically longer distance to parks in upper-income CBGs than to parks in middle-income and lower-income CBGs in both 2019 and 2020, with the exception of Community Park and Flagship Park in 2019.
The travel distance to parks located in upper-income level CBGs had the smallest percentage decrease (Figure 2-7c), which was -10.6% (95% CI: -10.9%, -10.3%), while it was -16.0% (95% CI: -16.3%, -15.7%) for parks located in middle-income CBGs, and was -14.1% (95% CI: -14.4%, -13.8%) for parks located in lower-income CBGs. (Table A-6)

Examining by park type, for Community Park, Flagship Park, Recreational Field/Courts and Triangle/Plaza park types, the parks located in upper-income level CBGs experienced the smallest percentage reduction in travel distance. While for Nature Area, Jointly Operated Playground and Playground park types, the parks located in lower income level CBGs had an increase in travel distance.

Figure 2-7 (a) and (b): Mean travel distance by park type and by income level of park CBGs; The letters to the right of the mean travel distances are the Tukey HSD multi-group comparison results between income groups for each park type, any common letter shared by two income groups indicates that the two groups were found to belong to the same group. (c): Percentage change of mean travel distance by park type and by income level of park CBGs, with 95% CI error bars.
2.4 Discussion

Change in the number of park visits and visitors.

The data show a sharp decrease in park visits and visitors following the start of New York City related COVID-19 pandemic restrictions in March 2020, which for most parks lasted throughout the year. Across the boroughs, we observed that Lower Manhattan experienced the largest decrease in park visits and visitors. As Lower Manhattan is NYC’s central area for business, culture, and government, it is to be expected that visitation to this area decreased substantially after the pandemic restrictions were imposed. In contrast, Staten Island, which is more suburban with fewer tourist attractions, experienced the smallest reduction in park visits and visitors. By park type, Triangle/Plaza and Flagship Park saw the biggest decrease in visits and visitors among all park types. Triangles/Plazas are smaller park areas and mostly located in densely populated business areas, the significant decrease in visits could be due to people not traveling to office buildings for work and therefore not using those Triangles/Plazas. Flagship Parks are destination attractions, even for NYC residents, and thus likely saw fewer visitations because people reduced destination-like leisure activities. Nature Areas had the smallest reduction in park visits and visitors, and even saw some increases above pre-pandemic level during summer months. Nature Areas are typically located on the outskirts of NYC or Staten Island (Figure 2-1), where concerns about crowding are reduced. The many services and benefits Nature Areas provide, from exercising, nature viewing, and birding, have also been shown to promote stress relief and mental health support during the pandemic. From a policy perspective, certain types of parks in NYC were ordered closed from April through June, including Playgrounds (Figure A-3) and Jointly Operated Playgrounds (jointly operated with the NYC Department of Education, Figure A-4). The biggest decrease in park visits did occur in
April 2020, however, the data showed a gradual rebound of visits to these parks while the order was in effect until summer months when temperatures may be too high for outdoor activities (Figure 2-3), indicating that compliance with the lockdown policy may have declined overtime. In addition, as per observed trends for different types of park visits and visitors are similar, there is no indication that NYC residents behaved differently in terms of park visitation than non-NYC residents.

**Change in park visits by NYC residents based on income level of park CBGs.**

The decision to classify parks by income level was based on the fact that, neighborhood income was found to be an influential factor that impacts park usage in previous studies.\textsuperscript{119-121} Furthermore, while many parks are designed to serve their immediate neighborhood, certain types with bigger sizes, such as Flagship Park, Community Park and Nature Area, could attract people from their larger service areas that may span the entire metropolitan region (Table A-7). This means that visitors may travel to parks in neighborhoods with different income levels than their own.

We observed that, parks located in higher-income neighborhoods experienced a smaller decrease in visits by NYC residents. This pattern was mainly driven by Flagship Park, Community Park and Nature Area, which are the three types of parks that usually have much larger areas (Figure 2-1, Figure A-2 and Table A-7); they saw greater reduction in visits in lower-income neighborhoods. Since park size was usually positively related with park use,\textsuperscript{122,123} and wealthier neighborhoods are usually considered safer,\textsuperscript{124-127} the combination of larger park areas, where social distancing is easier, located in wealthier neighborhoods likely explains this trend.
Recreational Field/Courts, which consist solely of hard surface and turf sports areas, showed a different trend, in that parks located in lower-income neighborhoods experienced a smaller decrease in visits by NYC residents. In many low-income communities, community located recreation facilities are often the only place for children to be physically active\textsuperscript{119}, which might explain greater usage of these facilities in these neighborhoods. For the rest of the types of parks, there was no significant difference in park visits change rate by NYC residents between income levels.

**Change in travel distance to parks for NYC residents.**

Through our analysis, we observed that mean travel distance to urban parks by NYC residents decreased from 5.9 km before the pandemic to 5.1 km during the pandemic. More specifically, there was a significant decrease in travel distance for all park types except Nature Area, Jointly Operated Playground, and Playground where travel distances remained similar to pre-pandemic levels. As previously discussed, Nature Area served as a safe haven for urban residents during the pandemic, with visits remaining relatively stable and even increasing in some months. People were willing to continue to travel to visit these areas despite the fact that they were frequently located on the fringes of the city\textsuperscript{26,103,128}. Jointly Operated Playground and Playground are scattered throughout the city's residential zones. Prior to the pandemic, travel distances to these two types of parks were the shortest among all types, implying that their primary role was to serve local residents. During the pandemic, visitors to them are still likely local residents, resulting in a small change in travel distance.

Visitors from lower-income CBGs tended to travel a statistically shorter distance to parks than those from middle-income and upper-income CBGs. This could be explained by the fact
that, people with higher incomes were willing to spend money on transportation and, as a result, were more able to travel to parks located a further distance away. Travel distances for visitors from upper-income CBGs decreased the most among the three income groups; this could be attributed to the fact that people with higher socio-economic status were more likely to have the ability to work from home.\textsuperscript{31,129,130}

When considering the income level of park CBGs, in general, people tended to travel a statistically longer distance to the parks located in upper-income level CBGs, and the travel distance to those parks decreased the least in percentage compared to parks in middle-income and lower-income CBGs. As high-income neighborhoods are usually perceived as a safer environment,\textsuperscript{124–127} this might explain people’s willingness to travel longer distances to the parks located there, even during the pandemic, resulting in a smaller reduction in travel distance.

First observations, the travel distance by NYC residents appear to be longer than expected, however, from other studies, we found comparable travel distance ranges. For example, one study conducted in Sapporo, Japan investigated the travel distances to urban parks and nature trails; they reported a mean travel distance to urban parks of 6.8-11.0 km before the pandemic and 4.9-10.6 km after the pandemic.\textsuperscript{131} Another study conducted in Wuhan, China found that the threshold travel distance (TTD, i.e., the third quartile travel distance for all visitors) to urban parks reduced from 4.2 km in pre-pandemic time to 3.0-3.9 km during different stages of the pandemic; TTD varied by travel modes, it ranged from about 1.1 km for walking to about 15 km for visitors taking buses and subways.\textsuperscript{132} The mean travel distance reported in our study reflected the general situation for all visitors with mixed travel modes, future studies can further disaggregate travel distances by travel modes when such information becomes available.
The importance of Nature Areas and Recreational Field/Courts.

Throughout the analysis from all these aspects, we would like to highlight Nature Areas, as they were the only type that saw a slight decrease or even some increases in visits and, furthermore, the travel distance to them also remained relatively stable. Perceived as a safe haven, these results demonstrated Nature Area’s critical role of serving urban residents in a time of crisis, and they should be well planned and maintained in the future urban developments. We would also like to highlight Recreation Field/Courts, particularly their significance in serving lower-income communities. They experienced a smaller decrease in visits than most types of parks except for Nature Areas, and they were the only park type that saw a greater decrease in upper-income CBGs and less decrease in lower-income CBGs. Travel distances to them have also fallen significantly, with the biggest decrease coming from people living in lower-income CBGs. These findings suggested that Recreation Field/Courts probably became the key destinations for local people during the pandemic, especially for those living in lower-income neighborhoods.

Limitations and avenues for future research.

Although we adjusted the park visits for temperature, which was a major factor in influencing park usage, other potential sources of bias in the data could exist. The number of devices in SafeGraph’s panel, for example, may change over time, introducing biases in visit counts; this issue could be especially prominent in cross-region studies\textsuperscript{108,133}. To overcome this bias, SafeGraph has recommended several methods for normalization\textsuperscript{134}. Since the supplementary datasets for normalization were not available at the time when we acquired the data for the analyses presented in this paper, we used the normalization by total visitors method.
to assess this potential bias in a secondary analysis. Overall, the normalized visits data were highly correlated with the raw visits data (Pearson’s R = 0.980, Table A-8). For individual park types, all park types showed a high correlation coefficient of over 0.9, except for Nature Area (Pearson’s R = 0.763). From further inspection, it appears that the raw visits data may have overestimated the visits to Nature Areas in 2019 or have underestimated the visits in 2020 (Figure A-5), implying the percentage increase in visits to Nature Area in 2020 could be even greater than reported here. While this does not violate our general conclusions, future research should consider this potential bias and make adjustments accordingly.

Our data showed that park neighborhood income level was a significant factor correlated with the park usage during the pandemic, however, it could be indicative of multiple factors that influence the choice to visit a park. Previous studies have found that parks in lower-income neighborhoods can be associated with poorer facilities, fewer services and less maintenance. Independent of neighborhood income level, park features such as sports facilities and water scenes, were found to be positively associated with park use; factors describing the surrounding environments of parks, such as population density, road density, distance to city center and accessibility via public transportation can also impact park usage. Future studies should investigate the underlying factors that are indicated by neighborhood income and influence park usage, therefore providing more detailed insights that could be used to improve park service.

Our findings indicate that there has been a fundamental change in park visitation habits of urban residents, and the change varies by park type and socio-economic status. Future studies should explore more in detail how other factors could potentially affect people’s park visitation,
such as age, gender, ethnicity and means of transportation, etc. Furthermore, the identified specific park types, such as Nature Area, for its importance to general urban residents; and Recreational Field/Court, especially for its importance to lower-income communities, should be well planned and managed to make urban parks a more resilient infrastructure system in the face of future crisis.
Chapter 3 – Quantifying the role of urban greening and other key factors in influencing the urban heat island effect

3.1 Introduction

The urban heat island (UHI) effect is a well-documented phenomenon whereby urban areas experience higher temperatures than their surrounding areas. This effect can be attributed to multiple factors associated with urban development, such as heat-retaining properties of construction materials, anthropogenic heat generation, reduced vegetation, and alternations in land surfaces, all of which foster heat retention and reduce cooling effects. The UHI effect exacerbates extreme heat events, posing significant health risks to urban populations, particularly during summer months. As urbanization continues to advance globally, mitigating the UHI effect has become a critical challenge for urban planners and policymakers.

Urban green spaces have been highlighted in the mitigation of the UHI effect, as they provide cooling opportunities in cities related to shading, the modification of the thermal properties of land surfaces and evapotranspiration, sometimes collectively referred to as the urban cooling island (UCI) effect. Many relevant studies have used remote sensing data to investigate how green spaces help in the cooling of urban environments, usually classifying the urban landscape into different land use/land cover (LULC) categories and examining their impact on land surface temperature (LST). The categorization of LULC ranges from a few general types such as built-up area, vegetation, bare land, water body, and so on, to a dozen detailed types. In these studies, the percentage coverage of each LULC type was usually
calculated at different spatial scales, then their relationships with LST were investigated for each LULC type to identify the most influential types. However, because a grid cell or land patch usually contains multiple LULC types, quantifying a grid cell or land patch in terms of a single numeric value of LULC can mask the relationship between the multiple features influencing LST in different settings.

The runoff coefficient, which quantifies the proportion of rainfall that becomes surface runoff and varies from 0 (no runoff generated) to 1 (all rainfall becomes runoff), is higher in hardscape areas with low water infiltration capacity and is lower in permeable, well-vegetated areas. By leveraging this variation, we propose an approach to quantify the impact of urban greening on LST, by using the runoff coefficient as an indicator of the degree of urban greening within a city landscape at different spatial scales.

Besides urban greening, other land surface properties, urban topography and morphology features can also be influential to LST. Higher albedo surfaces, for instance, can reflect more solar radiation back to the atmosphere, therefore reducing the heat absorption and leading to lower surface temperatures \(^{144-146}\), especially in dryer regions \(^{147}\). Land surface elevation is another factor that has been consistently reported to exhibit an inverse relationship with LST \(^{148-150}\). Water bodies such as lakes, rivers and bays, are known to exert a moderating effect on LST, often resulting in cooler daytime LSTs in their adjacent terrestrial areas \(^{151-154}\). The built environment further contributes to the spatial variability of LST; studies have identified a positive correlation between higher building density and higher LST \(^{155-157}\). Building height, on the other hand, has been reported to be negatively associated with daytime LST \(^{156-158}\).
suggesting that taller structures may contribute to lower daytime surface temperatures in urban areas possibly due to shading effects.

We focus on New York City (NYC), a densely populated urban area with diverse land cover types and features, as well as diverse socio-economic demographics. Using daytime LST in summer as a surrogate for the UHI effect, we examine the role of the runoff coefficient, along with other variables including surface albedo, distance to a water body, land surface elevation, building density and building height in influencing LST. Furthermore, we investigate disparities between different income groups and their exposure to extreme heat in order to shed light on the need for heat mitigation strategies for socially vulnerable communities.

Our study contributes to the growing body of research on the UHI effect by introducing the runoff coefficient as a quantitative indicator of urban greening and by also seeking to quantify the range of factors that might be responsible for controlling LST in a given urban grid cell or land patch. The findings from this study offer insights that can support urban land use planning to mitigate the UHI effect and enhance urban resilience to extreme heat events.
3.2 Data and Methods

Runoff Coefficient:

A high-resolution vector GIS dataset was created based on the NYC Planimetric Data\textsuperscript{159}, which represents a foundational geospatial basemap of New York City, encompassing digital orthophotography and detailed planimetric features derived from the high-resolution 2014 New York Statewide Flyover. The gaps in the planimetric map were filled using the NYC Land Cover Raster Data\textsuperscript{160}, which is a 6-inch-resolution 8-class land cover dataset generated from the 2017 Light Detection and Ranging (LiDAR) data acquisition. The processed dataset classified New York City into more than 13 million land patches of 100 land use types. To further improve its accuracy, manual inspection and correction were also undertaken. Runoff coefficient values were determined from a literature search\textsuperscript{161–163}, and were then assigned to each land use type (Figure 3-1). The major land use categories that account for more than 95% of the total NYC land area of 783.9km\textsuperscript{2}, as well as the runoff coefficients assigned to them are shown in Table 3-1. All water categories within the NYC boundary were assigned a runoff coefficient of 0.
Table 3-1 The major New York City land use categories (non-water type) and the associated runoff coefficients

<table>
<thead>
<tr>
<th>Land Use Category</th>
<th>Area (km²)</th>
<th>Area Percentage (non-water type)</th>
<th>Cumulative Area Percentage (non-water type)</th>
<th>Runoff Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building</td>
<td>159.6</td>
<td>20.36</td>
<td>20.36</td>
<td>0.95</td>
</tr>
<tr>
<td>Roadbed</td>
<td>124.5</td>
<td>15.88</td>
<td>36.24</td>
<td>0.85</td>
</tr>
<tr>
<td>Other Impervious</td>
<td>103.5</td>
<td>13.20</td>
<td>49.44</td>
<td>0.90</td>
</tr>
<tr>
<td>Tree Canopy</td>
<td>74.8</td>
<td>9.53</td>
<td>58.97</td>
<td>0.25</td>
</tr>
<tr>
<td>Grass/Shrubs</td>
<td>73.0</td>
<td>9.31</td>
<td>68.28</td>
<td>0.20</td>
</tr>
<tr>
<td>Park Boundary</td>
<td>70.9</td>
<td>9.04</td>
<td>77.32</td>
<td>0.20</td>
</tr>
<tr>
<td>Sidewalk</td>
<td>61.4</td>
<td>7.82</td>
<td>85.15</td>
<td>0.85</td>
</tr>
<tr>
<td>Parking Lots &gt; 2000 Square Feet</td>
<td>31.9</td>
<td>4.07</td>
<td>89.22</td>
<td>0.85</td>
</tr>
<tr>
<td>Cemetery Outline</td>
<td>15.8</td>
<td>2.02</td>
<td>91.24</td>
<td>0.20</td>
</tr>
<tr>
<td>Recreational area over 2 acres (Not NYC designated parks.)</td>
<td>8.8</td>
<td>1.12</td>
<td>92.36</td>
<td>0.70</td>
</tr>
<tr>
<td>Wetland</td>
<td>7.3</td>
<td>0.94</td>
<td>93.29</td>
<td>0.15</td>
</tr>
<tr>
<td>Bridge</td>
<td>6.3</td>
<td>0.80</td>
<td>94.10</td>
<td>0.85</td>
</tr>
<tr>
<td>Vacant Area containing no structures</td>
<td>4.8</td>
<td>0.62</td>
<td>94.71</td>
<td>0.30</td>
</tr>
<tr>
<td>Bare Soil</td>
<td>4.5</td>
<td>0.57</td>
<td>95.28</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Surface Temperature and Albedo:

To obtain surface temperature and albedo data, Landsat 8 OLI satellite images were downloaded from the USGS Earth Explorer data portal. Collection 2 Level 2 data were chosen since they were atmospherically corrected and treated as ready-to-use data. As we focused on the summer daytime land surface temperature, a total of 8 imagery sets taken from July 1st to August 31st in 2019 were obtained, which were all taken at around 15:30 local time. Since some land areas were covered by clouds in the images, quality assessment pixels were used as masks to
filter out those pixels with a mid to high possibility of being clouds, as determined by the corresponding pixel values indicated in the Landsat 8 documentation 164.

The land surface temperature was converted from the ST band (band 10) using conversion coefficients from the Landsat 8 documentation (Equation 3-1).

\[
LST = \alpha_{10} \times 0.00341802 + 149
\]  

(3-1)

Where \( \alpha_{10} \) is the Landsat 8 OLI ST band (band 10).

For albedo, we used Equation 3-2 to convert narrow band observations to broad band albedo 165,166.

\[
\alpha = 0.356\alpha_2 + 0.130\alpha_4 + 0.373\alpha_5 + 0.085\alpha_6 + 0.072\alpha_7 - 0.0018
\]  

(3-2)

Where \( \alpha \) is the converted broadband surface albedo, \( \alpha_2, \alpha_4, \alpha_5, \alpha_6, \alpha_7 \) are the corresponding Landsat 8 OLI bands 2,4,5,6,7 (blue, red, near-infrared and shortwave-infrared).

After the conversions, the average LST and albedo from the 8 images were calculated.

**Distance to Water Body:**

Different types of water bodies were classified in the land cover dataset, including 1) Ocean/Bay, 2) River, 3) Stream wider than 8 feet, 4) Lake/Resevoir, 5) Pond and 6) Swimming Pool. As we were interested in the potential cooling effects of the water bodies, we selected those water bodies with an area of at least 100 m\(^2\) to exclude the mixing effect of small water bodies. The shortest distance (edge to edge) between each pixel (30m x 30m) and the selected waterbodies was then calculated using the NEAR function in the ArcGIS Pro software.
Elevation:

The NYC 1 foot Digital Elevation Model (DEM) \(^{167}\) was obtained for analyses involving elevation. It provides a high-resolution, bare-earth elevation surface model for New York City, representing the ground’s elevation relative to sea level, with all above-ground features like buildings and vegetation removed.

Building Height:

A shapefile of footprint outlines of buildings in New York City obtained from the NYC Open Data \(^{168}\) was used to determine all building heights. This geospatial dataset provides detailed information about the physical characteristics of each building, including its geographical location, shape, and size. One of the key attributes in this dataset is the heightroof column, which represents the height of the building roof above the ground elevation. This attribute is measured in feet and provides an accurate representation of the vertical dimension of each building, it was used in this study to determine the height of all buildings.

Building Density and Per Capita Income:

The Census Block Group (CBG) boundaries data for NYC were obtained from the US Census Bureau \(^{109}\). CBG is one of the smallest geographical units used by the Census Bureau for reporting specific data, which typically contains between 600 and 3,000 people, making it particularly useful for capturing granular patterns. To provide a normalized building density metric, the total building areas in each CBG were summed up and were then divided by the corresponding CBG area. The resulted building density is presented as a ratio ranging from 0 to 1, offering a standardized metric that facilitates comparisons across CBGs of varying sizes.
The per capita income data at the CBG level for NYC were obtained from the American Community Survey (ACS) via the US Census Bureau data portal for the year of 2019. This dataset offers insights into the average income earned by individuals in a given CBG, serving as a valuable metric for understanding economic conditions and disparities across different regions.

**Analysis Methods:**

To investigate the relationship between LST vs runoff coefficient and other urban features/variables, namely albedo, distance to water body, elevation, building height, and building density (building area ratio), all variables were aggregated at the same resolution (30m × 30m). Ordinary Least Square (OLS) linear regression and XGBoost regression were then used to model the relationships between LST and the different variables.

For the OLS regression, we also used feature selection methods, including sequential forward selection (SFS) and sequential backward selection (SBS) to identify the most relevant set of features for predicting LST. SFS begins with no features and incrementally adds those that most enhance model performance, gauged by adjusted R² in our case, until no further significant improvement is observed. In contrast, SBS starts with all features, removing the least impactful ones iteratively. Both SFS and SBS were enhanced with a "floating" option, which allows for the addition or removal of features at any step, not just the current one. This flexibility can lead to a more optimal feature subset as it revisits previous decisions based on the current set of features, potentially improving model performance by considering a broader set of feature combinations.

The hyperparameters of the XGBoost regression model was optimized using the Hyperopt module, which employs a Bayesian optimization approach to select the best
combination of parameters from a predefined search space, aiming to improve the model's predictive performance. In addition, the relative importance and contribution of each feature variable to the model’s predictions were evaluated using the SHAP module. SHAP provides a game-theoretic approach to explain the output of any machine learning model, offering insights into which features are most influential and how they impact the model's output.

For detailed effects of specific variables on LST, as well as the exposure to LST by income group, the multi-group comparisons were conducted mainly using Tukey’s HSD post-hoc test at either the pixel level (30m × 30m) or the CBG level, which were specified in the following sections.
Figure 3-1 Maps of runoff coefficient and other land surface features in New York City
3.3 Results and Discussion

3.3.1 Relationship between LST and runoff coefficient and other urban features

3.3.1.1 OLS Linear Regression

As shown in Figure 3-2, LST is highly positively correlated with the runoff coefficient, with the runoff coefficient alone explaining 50.4% of the observed variations in LST. An increase of 0.1 in runoff coefficient will lead to an increase of 1.10 °C in LST.

To further explore the pixels with extreme temperatures, we divided the runoff coefficient into 50 equally spaced bins, then assigned 1) the pixels two standard deviations above the mean temperature in each bin as high temperature pixels, 2) the pixels two standard deviations below the mean temperature as low temperature pixels, and 3) the rest as intermediate temperature pixels. Then we used SFS and BFS methods as described in the “Data and Methods” section to evaluate the optimal feature subsets in predicting LST for the three temperature conditions and all data points, where the features included were runoff coefficient, distance to a water body, surface albedo, land surface elevation, building height and building density. For distance to a water body, the natural log of the value was used based on exploratory data analysis that determined how best to incorporate this feature into the regression models.
As shown in Table 3-2, different features were selected for predicting LST for each of the four conditions. The selected features were consistent by using both SFS and SBS methods, and all the selected features had a high significance level (p<0.001). Runoff coefficient was selected for all conditions, and it exhibited a positive correlation with LST. Distance to water also showed a positive correlation with LST, while it was not chosen in high temperature pixels probably because the high temperature pixels were away from the water bodies. Albedo and elevation both correlated negatively with LST, with pixels containing higher albedo surfaces (which would reflect more sunlight) and pixels with higher land surface elevations reducing LST, whereas building height showed a positive relationship with LST for high temperature pixels but a negative relationship for low temperature pixels. Building density, which was selected for low temperature pixels, showed a positive relationship with LST. As anticipated, higher adj. R² values are associated with increased features in the OLS regressions. Additionally, the role of each feature in LST prediction changes with the temperature condition.
Table 3-2 Selected feature subsets for predicting LST based on SFS and BFS feature selection methods

<table>
<thead>
<tr>
<th>Temperature condition</th>
<th>Features (p&lt;0.001 for all selected features)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>runoff coefficient</td>
</tr>
<tr>
<td>All</td>
<td>9.1715</td>
</tr>
<tr>
<td>High</td>
<td>8.2094</td>
</tr>
<tr>
<td>Low</td>
<td>10.0117</td>
</tr>
<tr>
<td>Intermediate</td>
<td>9.9375</td>
</tr>
</tbody>
</table>

3.3.1.2 XGBoost Regression

As shown in Figure 3-3a, the XGBoost regression using all six land cover features resulted in a model with a testing accuracy of 0.73 using the entire LST dataset.

![Figure 3-3 XGBoost regression results: (a) testing accuracy, (b) feature impact on model output, and (c) feature importance by mean absolute SHAP value](image)

The impact of each feature on model output is shown in Figure 3-3b. Runoff coefficient was shown to be consistently positively correlated with LST; building density (i.e. building area ratio) and distance to water body were also generally positively correlated with LST. Albedo and elevation were generally negatively correlated with LST; some low SHAP values (indicating low LSTs) associated with low albedos could be due to the evaporative cooling effect of the low-
albedo water surfaces\textsuperscript{144}, and those low SHAP values associated with low elevations could be attributed to the low elevation areas that are close to water bodies.

The relative feature importance was ranked by their SHAP values (Figure 3-3c), runoff coefficient was the most influential factor in predicting LST, followed by building density, distance to water body, albedo, elevation and building height.

### 3.3.2 Effect of surrounding runoff coefficients on LST

![Figure 3-4 CBG runoff coefficient of pixels in different temperature groups for (a) low runoff coefficient pixels, (b) medium runoff coefficient pixels and (c) high runoff coefficient pixels; means and medians were shown as triangles and orange lines.](image)

Pixels that have similar runoff coefficients still showed variations in LST (Figure 3-2), and this could be influenced by their surrounding environments. To investigate this, we used the three terciles (0.618, 0.826, 0.950) to divide runoff coefficient into three groups: low, medium and high. We then calculated CBG level runoff coefficients and attributed them to the pixels contained in the corresponding CBG as a surrogate for the pixels’ surrounding conditions. The
pixels were also characterized by their temperature condition as defined above and as shown in Figure 3-2.

As shown in Figure 3-4a,b, for low and medium runoff coefficient pixels, their corresponding CBG runoff coefficients showed a clear decreasing trend from high LST to low LST pixels, indicating the variation in their LSTs could be further attributed to their surrounding runoff coefficient conditions; for example, for two pixels with similar runoff coefficients, the one surrounded by land with a lower runoff coefficient (i.e., more green space) will have a lower LST than the pixel surrounded by land with a higher runoff coefficient. This trend does, however, began to diminish for high runoff coefficient pixels (Figure 3-4c), as CBG runoff coefficients for intermediate temperature pixels were slightly higher than those of high temperature pixels. Nonetheless, CBG runoff coefficients for low temperature pixels were still significantly lower than those of intermediate and high temperature pixels.

Overall, the results illustrated in Figure 3-4 highlight the influence not only of local runoff coefficients (and hence urban greening) on urban LST, but also the influence of the runoff coefficient of surrounding areas.
3.3.3 Effect of different types of water bodies on LST

To understand the difference of different water body types on LST, we compared the LST of the pixels immediately next to the water bodies (i.e., within 30m of the water bodies). Most types of water bodies showed a significant cooling effect except for Swimming Pools (Figure 3-5). Compared to the cityside mean LST of 38.39°C, Bay/Ocean showed the biggest cooling effect, with a mean LST of 32.41°C or a cooling degree (defined as degrees below the citywide cityside mean LST) of 5.98°C, followed by Lake/Reservoir, Pond, River, Stream wider than 8 feet and Swimming Pools, with mean LSTs of 32.74°C, 33.75°C, 34.09°C, 34.26°C and 38.31°C, or cooling degrees of 5.65°C, 4.64°C, 4.30°C, 4.13°C and 0.08°C, respectively.
Water bodies moderate urban LSTs through several mechanisms, such as evaporative cooling, thermal inertia, and wind and breeze effects. First, water absorbs heat from the environment when it evaporates, reducing the temperature of the surroundings; second, with a high specific heat capacity, water can absorb and store a large amount of heat without a significant increase in temperature; third, wind and breeze effects from water bodies, such as sea or lake effects, arises as cooler air over water replaces warmer rising air over heated land, thereby helping to cool adjacent urban areas. Furthermore, water bodies often support surrounding green spaces and vegetation, which provide additional cooling through shading and transpiration; this can further enhance the cooling effects of Lakes/Reservoirs that are usually located in parks or green spaces. The cooling effect of Swimming Pools was restricted likely due to their small size and location in residential areas.

3.3.4 Effect of land surface elevation on LST

Figure 3-6 LST of pixels in different elevation groups that were at least 90m away from water; means and medians were shown as triangles and orange lines.
In exploring the effect of land surface elevation on LST, we wanted to eliminate the potential confounding effect of water bodies on LST at low elevation waterfront areas. Thus, we excluded the pixels within a certain distance from water bodies and found, via exploratory data analysis, that a decreasing trend in LST vs land surface elevation started to show from about 90m away from a water body (a distance of 3 pixels). Using the three terciles of elevation, the pixels were divided into three elevation groups ranging from low to high.

As shown in Figure 3-6, LSTs in higher elevation groups are significantly lower than those in lower elevation groups. The mean LST of 38.06°C in high elevation group is slightly lower than the citywide mean LST of 38.39°C; while the mean LSTs in medium and low elevation groups are higher than that, which are 39.48°C and 39.59°C, respectively.
3.3.5 Effects of building features on LST

3.3.5.1 Effect of building density on LST

![Graph showing LST in different building density groups](image)

Figure 3-7 CBG LSTs in different building density groups; means and medians were shown as triangles and orange lines.

The trend of LST vs building density was investigated at the CBG level by dividing CBGs into three building density groups ranging from low to high, using the three CBG building density terciles, then the mean LSTs were calculated for each CBG building density tercile. As shown in Figure 3-7, the mean LSTs for low, medium and high building density CBGs were 38.88°C, 39.97°C and 39.93°C, respectively. Medium and high building density CBGs had significantly higher LSTs than low building density CBGs, while the LSTs in medium and high building density CBGs didn’t show a significant difference. This could be attributed to the fact that some high building density CBGs, such as the financial district, are close to big water bodies, therefore the warming effect of higher building density and the cooling effect of water bodies are mixed.
3.3.5.2 Effect of building roof albedos on LST

![Box plot showing LST of building roof pixels in different albedo groups](image)

Figure 3-8 LST of the building roof pixels in different albedo groups; means and medians were shown as triangles and orange lines.

When considering the entire urban land cover, the relationship between albedo and LST could be confounded by low albedo open water bodies, as their evaporative cooling effects could offset the heat absorption of the low albedo surface. Since cooling roofs with higher albedos have been implemented in many cities, including NYC, as an effort to mitigate excessive heat, we focused simply on data relevant to building roofs to assess the albedo effect on LST.

As illustrated in Figure 3-8, although LSTs for building roofs are all higher than the citywide mean LST, increased building roof albedos did result in significantly lower LSTs. The mean LSTs for building roofs in low, medium and high albedo groups were 43.41°C, 42.28°C and 41.80°C, respectively.
### 3.3.5.3 Effect of building height on LST in very high building density CBGs

In order to investigate the trend of LST vs building height in very dense built-up areas, or CBGs with very high building densities, we calculated the mean and standard deviation of building density (building area ratio) for all CBGs in NYC, which were 0.31 and 0.11, respectively. We then used the mean plus one standard deviation – 0.42 as the threshold to determine CBGs with very high building density. Then very high building density CBGs were divided into three building height groups from low, medium to tall using Jenks natural breaks classification method, which minimizes each class's average deviation from the class mean while maximizing inter-class differences. Business zones with high number of tall skyscrapers, such as midtown Manhattan, the Financial District and downtown Brooklyn, were successfully identified as tall building height groups (Figure 3-9a).

In very high building density CBGs, the LSTs in taller building height groups were significantly lower than those in lower building height groups, the mean LSTs for low, medium and tall building groups were 40.61°C, 39.66°C and 38.53°C, respectively (Figure 3-9b).
3.3.6 The unequal exposure to the heat across income groups

To investigate the exposure to urban heat across different income groups, CBGs in NYC were divided into lower-, middle- and upper-income groups using the three terciles of CBG per capita income, then the pixel LSTs in these three income groups were compared.

As shown in Figure 3-10, LSTs in the upper-income group were significantly lower than those in the middle- and lower-income groups. The mean LST of 37.96°C in the upper-income group was lower than the citywide mean LST of 38.39°C, while the mean LSTs in the middle- and lower-income groups were higher than that, at 39.06°C and 39.31°C, respectively.

Figure 3-10 again confirms the disproportionately higher exposure to urban heat in lower-income communities, as has been identified in many cities in the US and worldwide. To combat this social inequality, policy makers need to develop neighborhood-specific heat mitigation strategies to reduce excessive heat stress in socially vulnerable communities.
3.4 Conclusions

In this study, we used the runoff coefficient as an indicator of urban greening indicator and investigated its impact on daytime summer Land Surface Temperature (LST), along with five other land surface features and attributes, including surface albedo, distance to water bodies, land surface elevation, building density and building height within the context of New York City (NYC). For most of our analyses we compared LST with these urban features at a resolution of 30m×30m. We refer to these 30m×30m land areas as pixels.

From our OLS linear regression results, the runoff coefficient of a pixel was highly positively correlated with the LST of a pixel (p < 0.001), and itself can explain 50.4% of the observed variations in LST. According to the regression results, every 0.1 increase in runoff coefficient was associated with 1.10 °C increase in LST (95% CI: 1.10 – 1.11), indicating that urban greening can support reduced summertime land surface temperatures. The XGBoost regression model using all 6 land surface variables achieved a testing accuracy of 0.73; the feature importance evaluated by the mean absolute SHAP value indicated that runoff coefficient was the most important factor in predicting LST, followed by building density, distance to water body, albedo, land surface elevation and building height.

For pixels with similar runoff coefficients, we found that variations in LST could be further attributed to differences in the runoff coefficient of their surrounding environments, i.e., pixels surrounded by lower runoff coefficient surfaces may have lower LSTs than those surrounded by higher runoff coefficient surfaces. In terms of the water bodies examined, all 6 types, but the type Swimming Pool, showed a significant cooling effect on LST in adjacent pixels, with Bay/Ocean showing the strongest cooling effect, followed by Lake/Reservoir, Pond,
River and Stream wider than 8 feet. These 5 types of water bodies (excluding the swimming pools) provided cooling degrees ranging from 5.98°C to 4.13°C compared to the citywide mean LST. As for elevation, the decreasing trend of LST vs land surface elevation began to show when a pixel was at least 90m away from a water body, with the mean pixel LSTs in low and medium elevation groups being higher than the citywide mean LST, and the mean pixel LSTs in high elevation group being slightly lower than the citywide mean.

In terms of the building features analyzed, pixel LSTs in low building density CBGs were significantly lower than those in medium and high building density CBGs, with the latter two groups showing no significant difference between pixel LSTs. Additionally, the mean LSTs in all three CBG building density groups were higher than the citywide mean. For building roofs, building roofs with higher albedos did have significantly lower LSTs than those in medium and low albedo groups. Nonetheless, the mean LST for all building roofs, regardless of whether they were within a high, medium or low albedo group, were higher than the citywide mean. In very high building density CBGs, the LST of pixels located in taller building height groups, versus medium or low building height groups, showed significantly lower LSTs. Again, however, the mean LST of pixels located in very high building density CBG was still always above the citywide mean.

In assessing the exposure to urban heat across different income groups, we found that LSTs in upper-income communities were lower than the citywide mean, while lower-income communities were exposed to significantly higher LSTs. The heat mitigation strategies can be developed by modifying specific or a combination of land surface features investigated in this
study, while lower-income communities should be given more attention in these developments to protect socially vulnerable populations.
Chapter 4 – A LoRaWAN-based environmental sensing network for urban green space monitoring with demonstrated application for stormwater management

4.1 Introduction

Urban green spaces play a crucial role in enhancing the ecological and social fabric of cities. Specifically, these spaces can provide a myriad of benefits, from mitigating urban heat island effects to improving air quality to providing recreational opportunities for urban residents. However, the health and sustainability of these green spaces are integrally linked to their soil conditions and health. In particular their soil moisture dynamics are key to plant growth, microclimate regulation, and stormwater management, among other important factors. Improving understanding of the dynamics of soil moisture in urban green spaces, especially during wet weather conditions, is thus an important and essential step for effective urban green space management and the development of design and implementation strategies for resilient urban ecosystems that can help reduce flooding during wet weather flow.

In recent years, advances in environmental sensing technologies have opened new avenues for monitoring soil conditions in urban landscapes. Traditional soil condition measurement methods, such as lab tests and in-situ measurements using various sensing instruments, while reliable, often involve complex sampling and/or installation protocols and high costs, making them unfavorable for large-scale deployment in diverse urban settings. Wireless sensing technologies, such as Low Power Wide Area Networks (LPWAN), offer a promising alternative because of their ability to offer affordable connectivity to low-power
devices distributed over large geographical areas. Among these, Long Range Wide Area Network (LoRaWAN) has emerged as a popular choice due to its low power consumption, long-range communication capabilities, and ease of deployment.\(^ {88,89}\) LoRaWAN-based systems have been used in multiple aspects of environmental monitoring, including air quality\(^ {90,91}\), weather conditions\(^ {92}\), tree health\(^ {93}\), and soil conditions\(^ {94}\), etc. However, most of these studies only tested their LoRaWAN system for a short period of time, ranging from a few days to several weeks, and just reported the raw observations of the measured parameters, with no additional analysis and modeling development, thus providing limited demonstration of the actual utility of the system for advancing understanding and supporting decision making. Additionally, no LoRaWAN-based soil sensing network has been developed and tested in diverse urban settings.

In this study, we developed a LoRaWAN-based environmental sensing system which measures air temperature, air humidity, soil temperature and soil moisture, with lab calibration to enhance its reliability. The system was deployed as a network across seven different urban green spaces on an urban campus, including tree pits, lawns, and shrub settings, etc. Additionally, meteorological data was collected from a local weather station. Over the course of one year, this network provided continuous monitoring of the aforementioned environmental parameters, enabling us to capture and analyze the trend of soil moisture in different green spaces and its response to various meteorological conditions, including rainfall events. We used this information to identify factors influencing temporal trends in soil moisture across the different monitored green spaces. To demonstrate the utility of the network, we also established quantitative relationships between the change in soil moisture during a rainfall event based on the rainfall amount and initial soil moisture level preceding rainfall, as one example of how such
a network might be used to understand the role of different green spaces in reducing stormwater runoff during wet weather flow.

Our approach underscores the potential of wireless sensing networks in urban ecological monitoring. Furthermore, data gathered from the network highlights the heterogeneity in soil moisture dynamics across different urban green spaces, providing valuable insight into urban green space management, particularly in the context of stormwater management. Such insight can support city planners and environmental managers in decision making that optimizes the ecological benefits of urban green spaces and enhances urban resilience in the face of stressors, such as changing precipitation patterns due to climate change.
4.2 System Design and Methods

4.2.1 Sensing system design

4.2.1.1 Major components

The major components of the sensing system are listed in Table 4-1. The Arduino MKR WAN 1300/1310 board enables long range communication using LoRa wireless protocol or through LoRaWAN networks, with sensors communicating with the microcontroller through their respective interfaces. A circuit board was designed to connect all these devices, and the system was powered by 8×rechargeable AA batteries using a 3.3V buck converter to regulate the voltage. A weather-proof sensor box was also designed and 3D printed to house the system in the field (Figure 4-1b).

<table>
<thead>
<tr>
<th>Component</th>
<th>Model/Type</th>
<th>Accuracy</th>
<th>Operating Conditions</th>
<th>Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microcontroller</td>
<td>Arduino MKR WAN 1300/1310</td>
<td>\</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air temperature/humidity sensor</td>
<td>Adadruit SHT-30 Mesh-Protected Weather-Proof Temperature/Humidity Sensor</td>
<td>±2% relative humidity, ±0.5°C</td>
<td></td>
<td>I°C</td>
</tr>
<tr>
<td>Soil temperature sensor</td>
<td>Waterproof 1-Wire DS18B20 Compatible Digital Temperature Sensor</td>
<td>±0.5°C</td>
<td>-10°C to +85°C</td>
<td>1-Wire</td>
</tr>
<tr>
<td>Soil moisture sensor</td>
<td>Vegetronix VH-400 Soil Moisture Sensor</td>
<td>2% (at 25°C)</td>
<td>-40°C to +85°C</td>
<td>Analog</td>
</tr>
</tbody>
</table>
Figure 4-1 (a) System diagram and illustration of the four different vegetation settings used in study, (b) 3d-printed sensor box and system assembly, (c) LoRaWAN gateway, (d) weather station, (e) system in operation in the field (node 7), (f) real-time data monitoring dashboard
4.2.1.2 Auxiliary equipment

There are two important pieces of auxiliary equipment: 1) a high performance LoRaWAN gateway designed and manufactured by the Columbia University Internet Real-Time (IRT) Lab, which provides a full campus signal coverage and beyond (Figure B-1); 2) a Davis Instrument 6152 Vantage Pro2 Wireless Weather Station, which measures a range of weather parameters every 15 minutes, such as air temperature, air humidity, atmospheric pressure and rain rate, etc. The two devices were both installed on the roof of the S.W. Mudd Building located on the northeast corner of Columbia University’s Morningside Campus to avoid obstructions (Figure 4-1c,d).

The sensing systems were registered and configured on The Things Network (TTN)\(^{185}\), allowing the data to be forwarded from the sensing systems to our local server and database. We also developed a data dashboard using Grafana to enable real-time data monitoring (Figure 4-1f).

4.2.2 Soil moisture sensor calibration

The Vegetronix VH-400 soil moisture sensor outputs raw voltage values. Although the manufacturer provided an approximate piecewise curve to convert sensor output voltage to volumetric water content (VWC), we observed that voltages above 2.2V did not accurately represent known volumetric water contents. Thus we conducted a lab calibration to obtain a more accurate piecewise conversion curve.

The standard soil gravimetric water content (GWC) measurement procedure was used for the calibration\(^{84,186}\): 1) soil samples were collected from a measurement site and dried in an oven with a temperature over 105°C for 24 hours; 2) water was added to the dried soil sample to achieve a moisture level, and the sample was then stirred evenly using a mixer; 3) the soil
moisture sensor was placed into the soil sample to record a voltage reading, then the wet soil sample was weighted \((m_{\text{wet}})\); 4) the soil sample was dried in the oven again for 24 hours, and its dry weight \((m_{\text{dry}})\) was recorded; Equation 4-1 was then used to calculate the soil GWC.

\[
GWC = \frac{m_{\text{wet}} - m_{\text{dry}}}{m_{\text{dry}}} \tag{4-1}
\]

We tested 3 soil moisture sensors to verify their consistency. A total of 6 moisture levels were measured, including a fully dried condition and a fully saturated condition; at each moisture level, at least 10 voltage readings were taken for each moisture sensor. The results confirmed that the readings from the 3 sensors were highly positively correlated with each other (all Pearson’s Rs > 0.99, p < 0.001).

To convert GWC to VWC, soil bulk density (BD) needed to be measured. We collected 3 relatively undisturbed soil samples from a measurement site using cylinder soil samplers with known volumes \((V_s)\) (Figure 4-2a), then dried the soil samples in the oven for 24 hours and measured their dry weight \((m_{\text{dry}})\). Equations 4-2 and 4-3 were then used to determine BD and VWC.

\[
BD = \frac{m_{\text{dry}}}{V_s} \tag{4-2}
\]

\[
VWC = GWC \times BD \tag{4-3}
\]

Following these conversions, the calibration curve of VWC vs sensor voltage output was modeled into both an overall linear curve \((R^2=0.963)\) and a piecewise linear curve using ordinary least square (OLS) linear regression. The aggregated calibration results were shown in Figure
4-2b, we used the piecewise linear curve to calculate VWC from raw voltages for the analyses in the following sections.

![Figure 4-2](image)

Figure 4-2 (a) Collected soil samples, (b) soil moisture sensor calibration curves, (c) soil texture estimation by sedimentation method (node 4)

### 4.2.3 Sensor deployment and sites description

The sensor box was usually attached to a tree trunk at or near the measurement site, with the air temperature/humidity sensor also attached to it at a height of 1.5-2.0m above ground; the soil moisture sensor and the soil temperature sensor were buried in the soil at a depth of 15-20cm, with the two sensors being about 0.5-1m away from each other horizontally.

The sensor systems were deployed at seven green spaces across the Columbia University Morningside Campus, referred to as nodes 1 to 7, to constitute a sensing network (Figure 4-3), the data collection was conducted from May 7th, 2022 to May 7th, 2023 to cover a full calendar year.
Soil samples were collected from each measurement site, and the soil textures were estimated using the sedimentation method, which is based on the principle that different soil particles settle in water at different rates due to their varying sizes and densities (Figure 4-2c). Sand particles, being the heaviest and largest, settle first usually within a few minutes of a sedimentation test; following that, the smaller and lighter silt particles begin to settle over several hours; and the smallest and lightest particles, clay, take the longest to settle, often up to several days. Once all the particles have settled and the proportions of sand, silt and clay have been identified, the soil texture can be determined using the soil texture triangle. The sites description, along with the estimated soil textures, are shown in Table 4-2.
Table 4-2 Sites description and soil texture estimations for each site

<table>
<thead>
<tr>
<th>Site</th>
<th>Site Description</th>
<th>Soil Particle Proportions (Sand / Silt / Clay)</th>
<th>Estimated Soil Texture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node 1</td>
<td>A standalone tree pit site under an old northern red oak tree, with lawn in the surrounding area</td>
<td>52% / 45% / 3%</td>
<td>Loam</td>
</tr>
<tr>
<td>Node 2</td>
<td>A plant site with seasonal Hosta sieboldiana</td>
<td>50% / 43% / 7%</td>
<td>Loam</td>
</tr>
<tr>
<td>Node 3</td>
<td>A standalone tree pit site under an adult little-leaf linden tree, with lawn in the surrounding area</td>
<td>41% / 54% / 5%</td>
<td>Silt Loam</td>
</tr>
<tr>
<td>Node 4</td>
<td>A tree pit site under a young little-leaf linden tree, with evergreen Pachysandra terminalis (Japanese spurge) and older linden trees in the surrounding area</td>
<td>60% / 25% / 15%</td>
<td>Sandy Loam</td>
</tr>
<tr>
<td>Node 5</td>
<td>A grass/lawn site</td>
<td>14% / 77% / 9%</td>
<td>Silt Loam</td>
</tr>
<tr>
<td>Node 6</td>
<td>A grass/lawn site</td>
<td>37% / 61% / 2%</td>
<td>Silt Loam</td>
</tr>
<tr>
<td>Node 7</td>
<td>A tree pit site under a young little-leaf linden tree, with evergreen Pachysandra terminalis (Japanese spurge) and older linden trees in the surrounding area</td>
<td>44% / 41% / 15%</td>
<td>Loam</td>
</tr>
</tbody>
</table>
4.3 Results and Discussion

4.3.1 Distribution of soil moisture and the seasonal variation

Figure 4-4 illustrates the variation in soil moisture ($VWC$ or $\theta$) at each site or node across the entire monitoring period. As shown, variability existed at each site and across the sites. Node 4 had the lowest mean soil moisture of 0.27, although it had a similar physical setting as node 7, the soil at node 4 was a sandy loam soil type, which has a lower field capacity and lower water holding capacity than loam and silt loam found at the other sites. The grass/ lawn settings of node 5 and node 6, on the other hand, had the highest mean soil moisture of 0.57 and 0.60, respectively; aside from their silt loam soil type having a higher water holding capacity, there were concrete vaults beneath these two lawn sites which would inhibit soil drainage, and the soil depths were shallow (30-40cm), so could be saturated with small volumes of water. Other studies have also reported grassland soils being wetter than soils in forestland, shrubland or individual tree pits, which could be attributed to lower rainfall interception losses at grassland sites as well as lower transpiration losses from vegetations.
To examine seasonal variations in soil moisture at each of the monitoring sites, we divided the year into four seasons: spring (March, April and May), summer (June, July, and August), autumn (September, October, and November), and winter (December, January and February). As shown in Figure 4-5, node 1, node 3 and node 7 showed a similar trend whereby the soil moisture in summer was the lower than in any other season. New York City is characterized to have a humid subtropical climate according to the Köppen-Geiger climate classification\(^\text{192}\), summers in this climate are typically hot and humid. Furthermore, while occasional thunderstorms can provide significant rainfall, the high temperatures often lead to increased evaporation and transpiration, which can deplete soil moisture\(^\text{193}\). In densely built-up areas, the urban heat island effect can further exacerbate moisture loss from the soil\(^\text{194}\). On the other hand, node 2, node 4, node 5 and node 6 had their highest mean soil moistures in summer.
compared to other seasons. We attribute this to more frequent irrigation at these sites than the others during summer months.

![Figure 4-5 Distribution of soil moisture at each measurement site in different seasons](image)

**4.3.2 Increase in soil moisture vs rainfall amount and initial soil moisture**

To normalize soil moisture of each node to a common scale, soil moisture was converted to degree of saturation (DoS) using Equation 4-4, which was also called normalized water content or effective saturation 195.

\[
\text{DoS} = \frac{\theta - \theta_r}{\theta_s - \theta_r} \tag{4-4}
\]

Where \(\theta\) is soil moisture (in terms of VWC), \(\theta_r\) is residual soil moisture, \(\theta_s\) is saturated soil moisture. \(\theta_r\) is usually between 0.001 to 0.1 196,197, we assumed a universal \(\theta_r\) value of 0.05 for all nodes, and used the highest \(\theta\) we recorded for each node as their respective \(\theta_s\).
To model the response of soil moisture to rainfall events, we separated the monitored rainfall into individual rainfall events by a 6-hour minimum dry period, following the National Oceanic and Atmospheric Administration (NOAA) storm data preparation guidelines. Soil moisture responses to rainfall are influenced by multiple factors, such as storm depths and intensities, antecedent conditions, vegetation characteristics, soil properties, as well as topographical features and other site specifications.

To characterize the soil moisture response to rainfall, we defined the increase in soil degree of saturation ($DoS_{\text{increase}}$) in individual rainfall events as the difference between the highest degree of saturation recorded during the rainfall event and the initial degree of soil saturation ($DoS_{\text{initial}}$), which was the DoS 1 hour before the start of a rainfall event. In exploratory data analysis, we tested the relationship between $DoS_{\text{increase}}$ and $rain\_amount$ using both ordinary least squares (OLS) linear regression and Spearman’s rank correlation test. In OLS linear regression, we omitted the intercept term to enforce the physical constraint that, when $rain\_amount$ is 0, there should be no increase in soil moisture, i.e., $DoS_{\text{increase}} = 0$.

As shown in Figure 4-6, with the exception of node 5, all Spearman’s rank correlation coefficients ($\rho$) were positive and significant, suggesting that $DoS_{\text{increase}}$ increases monotonically with $rain\_amount$. Node 5 was almost always close to saturation, therefore there was little space for it to increase moisture. The results of the OLS linear regression also indicated a positive correlation between $DoS_{\text{increase}}$ and rainfall amount ($rain\_amount$); however, the adjusted $R^2$ values for several nodes were low. We further observed that $DoS_{\text{initial}}$ appeared to have an impact on $DoS_{\text{increase}}$: for example, if $DoS_{\text{initial}}$ was low, $DoS_{\text{increase}}$ could be larger with the same $rain\_amount$ than if $DoS_{\text{initial}}$ was high.
We then modeled $\text{DoS}_{\text{increase}}$ using Equation 4-5, which considered the interaction effect of $\text{rain\_amount}$ and $\text{DoS}_{\text{initial}}$, while again enforcing the physical constraint that there should be no increase in soil moisture when $\text{rain\_amount}$ is 0.

$$\text{DoS}_{\text{increase}} = \alpha \cdot \text{rain\_amount} + \beta \cdot \text{rain\_amount} \times \text{DoS}_{\text{initial}}$$ (4-5)

The model was fitted for each node using ordinary least square (OLS) regression, as shown in Figure 4-7, which provides information on parameters used to model behavior at each node. All nodes but node 5 achieved an adjusted $R^2$ of over 0.65. Since DoS is a normalized measure of soil moisture, we also combined data from all nodes and fitted an overall model, which achieved an adjusted $R^2$ of 0.642, also shown in Figure 4-7.

Surface runoff is formed when soil moisture is close to saturation, i.e., DoS is close to 1. While $\text{DoS} = \text{DoS}_{\text{initial}} + \text{DoS}_{\text{increase}}$, assuming that the DoS threshold value for runoff generation is $\text{DoS}_{\text{max}}$, the model described by $\text{DoS}_{\text{increase}} = \alpha \cdot \text{rain\_amount} + \beta \cdot$
rain_amount × DoS_{initial} \quad (4-5 \text{ can be translated to a model to determine runoff generation in two ways: 1) given } rain_amount, \text{ to determine the maximum initial DoS (DoS}_{initial,max}) \text{ that will cause runoff if exceeded, as formulated in Equation 4-6; 2) given DoS}_{initial}, \text{ to determine the maximum rainfall amount (rain_amount}_{max}) \text{ that will cause runoff if exceeded, as formulated in Equation 4-7.}

\[
DoS_{initial,max} = \frac{DoS_{max} - \alpha \cdot rain_amount}{1 + \beta \cdot rain_amount} \quad (4-6)
\]

\[
rain_amount_{max} = \frac{DoS_{max} - DoS_{initial}}{\alpha + \beta \cdot DoS_{initial}} \quad (4-7)
\]

From exploratory data analysis, we observed that when DoS_{initial} was 0.9 and above, there was almost no increase in DoS, we therefore used 0.9 as DoS_{max}. The resulted runoff threshold lines are shown in Figure 6. In general, nodes 1, 3 and 4 had a greater rainfall holding capacity in terms of runoff generation, while nodes 2, 6, and 7 were more prone to generate runoff. Node 6 was especially prone to runoff generation at almost all storm events above about 10mm.
Figure 4.7 Increase in soil DoS as a response to rainfall amount and initial soil DoS; the star marks after the regression coefficients indicate the significance level (***/p<0.001, **/p<0.01, */p<0.05)

Furthermore, we categorized rainfall events into different groups and calculated characteristic runoff coefficients for each node. Characteristic runoff coefficient was defined as the proportion of runoff to rainfall amount, and the runoff was calculated using Equation 4-8:

\[
\text{runoff} = \begin{cases} 
\text{rain\_amount} & (\text{DoS}_{\text{initial}} \geq \text{DoS}_{\text{max}}) \\
\max(0, \text{rain\_amount} - \text{rain\_amount}_{\text{max}}) & (\text{DoS}_{\text{initial}} < \text{DoS}_{\text{max}})
\end{cases}
\] (4-8)

Where \(\text{rain\_amount}_{\text{max}}\) was determined using Equation 4-7.

The runoff coefficients of different types of green spaces typically range from 0.15 to 0.25 in government guidelines for stormwater management and drainage systems design\textsuperscript{161,162}. As shown in Table 4-3, the overall runoff coefficients of nodes 1, 2 and 3 were comparable to this range, whilst node 4 was below this range and nodes 5, 6 and 7 were above it. Other patterns observed: 1) the runoff coefficient was lower when the rainfall amount was lower and became higher with increasing rainfall amount, this pattern was common across all nodes; 2) when the rainfall amount was above 30mm, runoff coefficients could be much higher than the typical...
range; 3) node 4, because of its large stormwater holding capacity and usually low $D_{S_{initial}}$ (Figure 5), was still able to have a low runoff coefficient even when rainfall amount was above 50mm; 4) in contrast, node 5 and 6, because of their usually high or near saturation $D_{S_{initial}}$, had high runoff coefficients regardless of the rainfall amount.

### Table 4-3 Characteristic runoff coefficient by rain group

<table>
<thead>
<tr>
<th>Rain Group (mm)</th>
<th>Number of Events</th>
<th>Total Rainfall Amount (mm)</th>
<th>Characteristic Runoff Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Node 1</td>
</tr>
<tr>
<td>0-10</td>
<td>68</td>
<td>197.9</td>
<td>0.002</td>
</tr>
<tr>
<td>10-20</td>
<td>19</td>
<td>273.6</td>
<td>0.000</td>
</tr>
<tr>
<td>20-30</td>
<td>12</td>
<td>293.0</td>
<td>0.101</td>
</tr>
<tr>
<td>30-50</td>
<td>5</td>
<td>192.2</td>
<td>0.534</td>
</tr>
<tr>
<td>50+</td>
<td>5</td>
<td>343.9</td>
<td>0.402</td>
</tr>
<tr>
<td>overall</td>
<td>109</td>
<td>1300.6</td>
<td>0.214</td>
</tr>
</tbody>
</table>

A direct implication of these findings could be, while green spaces can help retain and drain storm water, their performance varies by site specifications such as vegetation type, soil conditions and physical settings. Besides, more proactive soil drainage systems are needed to enable green spaces to reduce runoff during larger storm events which could have a rainfall amount of over 50mm, especially when more of these extreme events are expected under the changing climate.$^{83,204,205}$.
4.4 Conclusions

In this study, we developed a LoRaWAN-based environmental sensing system which measures air temperature/humidity, soil temperature and soil moisture. A LoRaWAN gateway was also established to provide extensive signal coverage, as well as a weather station to record meteorological data. After lab calibration, the sensing system was deployed as a sensing network covering seven different urban green spaces on Columbia University’s Morningside Campus, including tree pits, lawns and shrubs, etc. A data dashboard was developed to enable real-time data visualization of the monitoring results. The data collection from the monitoring system was conducted for a full calendar year from May 7th, 2022 to May 7th, 2023, at the same time demonstrating the reliability of the sensing network under different weather conditions.

From the collected data, we investigated the soil water dynamics in different green spaces. The observed soil moisture varied by vegetation types, soil conditions and physical settings, as well as seasonally; the mean soil moisture at monitored lawn sites was the highest among all types of green spaces, this could be due to 1) the silt loam soil type at these sites having higher water holding capacity, 2) the concrete vaults beneath and the shallow soil depth leaving little space for drainage, and 3) less interception and transpiration loss compared to other vegetations. On the other hand, node 4, while with a similar physical setting as node 7, had the lowest mean soil moisture, which could be caused by the low field capacity and low water holding capacity of its sandy loam soil type. For seasonal variations, three nodes showed their lowest soil moisture in summer, possibly due to the high temperatures that enhanced evaporation and transpiration. While the other four nodes, which received more frequent irrigations, showed high soil moisture in summer.
Furthermore, we modeled the soil moisture response to rainfall events by quantifying the increase in soil moisture using rainfall amount and initial soil moisture. This methodology yielded robust models for each monitored site, with adjusted R² values exceeding 0.65 for all but one location, demonstrating the effectiveness of our approach in capturing the complex interplay between rainfall and soil moisture in urban environments. After classifying rainfall events into different rainfall amount groups, we also calculated the characteristic runoff coefficient for each node using the developed model, and saw different performance of these green spaces in terms of runoff generation. With different configurations, green spaces can help retain and drain storm water from various sized rainfall events up to 30-50mm; however, for those extreme storm events with rainfall amounts of over 50mm, more proactive drainage systems are needed to prevent excessive runoff and flooding, especially when these events are projected under the changing climate.

This study contributes to the growing body of knowledge on urban environmental monitoring by demonstrating the practical application of LoRaWAN technology in real-world settings. Our findings provide valuable insights into the management of urban green spaces, particularly in the context of stormwater management. By understanding the soil moisture dynamics, city planners and environmental managers can make informed decisions to optimize the ecological benefits of urban green spaces, improve urban water management strategies, and enhance urban resilience in the face of climate change challenges.
Chapter 5 – Conclusions

5.1 Key Findings

In this dissertation, the multifaceted nature of urban sustainability challenges was explored through three distinct yet interconnected studies, all of which considered the role of urban green space and urban greening on supporting progress toward urban sustainability. The first study delved into the change in urban park usage by urban residents during the COVID-19 pandemic, revealing a significant shift in visitation patterns influenced by park type, location, and socioeconomic factors. The second study focused on the urban heat island effect, employing the runoff coefficient as an urban greening indicator to understand its influence on land surface temperature. The third project pioneered the use of a LoRaWAN-based environmental sensing system to monitor soil and other conditions across different green spaces, offering insights into soil moisture dynamics in urban green spaces and their crucial role in stormwater management. Together, these studies not only shed light on the complexities of urban environmental dynamics, but also provided valuable visions and knowledge relevant to advancing sustainable urban development and resilience in the face of evolving global challenges.

5.1.1 The importance of urban parks in a time of crisis

In Chapter 2, the extensive analysis of changes in urban park usage by urban residents during the COVID-19 pandemic in New York City provides several key findings:

1) **Overall decrease in park visits and visitors:** There was a marked decline in park visits and visitors across NYC from the year 2019 to 2020, particularly after pandemic restrictions began in March 2020. Lower Manhattan, a central area for
business, culture, and government, witnessed the largest decrease, while Staten Island, being more suburban, saw the smallest reduction.

2) **Variation by park type**: Different types of parks experienced varying rates of visitation change. Triangle/Plaza and Flagship Parks, often in densely populated business areas, saw the biggest visitation decreases in 2020 compared to 2019. Nature Areas, typically on the city's outskirts, had the smallest reduction and even saw increased visits in 2020 compared to 2019 during summer months.

3) **Impact of park location and socio-economic factors**: Parks in higher-income neighborhoods experienced a smaller decrease in visits in 2020 compared to 2019, particularly larger park types like Flagship Park, Community Park, and Nature Area. In contrast, Recreation Field/Courts in lower-income neighborhoods saw a smaller decrease in visits, suggesting their importance as local recreational spaces in these communities.

4) **Changes in travel distance**: The average travel distance to parks for NYC residents decreased during the pandemic. However, travel distances to Nature Areas and local parks like Playgrounds remained relatively stable, indicating their continued importance. Visitors from lower-income communities traveled shorter distances to parks, possibly due to limited transportation options.

5) **Significance of Nature Areas and Recreation Field/Courts**: Nature Areas were crucial during the pandemic, maintaining stable visitation and travel distances. Recreation Field/Courts were particularly significant in lower-income communities, with smaller decreases in visits and travel distances, highlighting their role as key local destinations during the pandemic.
These findings underscore the importance of different urban park types to urban residents, particularly during a public health crisis, and provide valuable insights for urban planning and the development of resilient public spaces.

5.1.2 The role of urban greening in mitigating excessive urban heat

In Chapter 3, the use of the runoff coefficient as an indicator of urban greening and its impact on Land Surface Temperature (LST) as an indicator of the urban heat island effect was explored, alongside five other land surface and urban features including surface albedo, distance to water bodies, land surface elevation, building density, and building height. New York City was used as a case study and relationships between LST and the six factors were explored at a pixel scale of 30m×30m. The key findings from this work are as follows:

1) **Runoff coefficient and LST relationship**: The runoff coefficient showed a strong positive correlation with LST. A 0.1 increase in the runoff coefficient corresponded to a 1.10 °C increase in summer daytime LST in NYC, indicating that more urban greening can lead to cooler surface temperatures. This single factor accounted for 50.4% of the variation in LST in NYC.

2) **XGBoost regression model**: Using all six urban features/variables, the model achieved a testing accuracy of 0.73. The runoff coefficient was the most significant predictor of LST, followed by building density and distance to water bodies.

3) **Influence of surrounding environments**: Pixel-level LST variations could be further explained by the runoff coefficient of their surrounding areas. For pixels with similar runoff coefficients, those surrounded by lower runoff coefficient surfaces (an
indicator of greater greenery) can have lower LSTs than those surrounded by higher runoff coefficient surfaces (an indicator of greater hardscape).

4) **Impact of water bodies and elevation**: All types of water bodies, except swimming pools, showed a significant cooling effect, with bay/ocean areas providing the most substantial reduction in LST. LST tended to decrease with increasing land surface elevation, especially when more than 90m away from water bodies.

5) **Building features and LST**: Lower building density areas had significantly lower LSTs compared to medium and high-density areas. Higher building roof albedos, as well as taller building heights in very high-building density areas, also led to lower LSTs.

6) **Socio-economic disparities**: Upper-income communities experienced lower LSTs than the citywide mean, while lower-income communities faced higher LSTs, highlighting the need for targeted heat mitigation strategies in these vulnerable areas.

The findings associated with this study emphasize the critical role of urban greening and several other urban/land surface features in urban heat mitigation, and underscore the importance of incorporating these factors in strategies to mitigate urban heat island effects, especially for socially vulnerable populations.

**5.1.3 Environmental sensing to support urban stormwater management**

Chapter 4 presents the development and implementation of a LoRaWAN-based environmental sensing system that was used to monitor several environmental parameters and soil water dynamics across seven diverse urban green spaces over a full year on Columbia
University’s Morningside Campus. Data collected from the monitoring system was used to analyze and model soil moisture responses to rainfall events at the seven different green spaces. Key findings from this particular study include:

1) **Soil moisture variability**: Soil moisture levels varied significantly across the seven different green spaces, and were influenced by vegetation types, soil types and conditions, physical settings, and seasonal changes. The monitored lawns generally exhibited the highest soil moisture, which was attributed to their silt loam soil type, underlying concrete structures, and reduced rainfall interception and transpiration loss compared to other green space types. In contrast, the only monitored site with a higher permeability sandy loam soil, showed the lowest moisture levels during the one-year monitoring period.

2) **Seasonal variations**: Some of the monitored green spaces experienced their lowest soil moisture levels in summer, likely due to increased evaporation and transpiration under the higher summer temperatures. Conversely, other green spaces that were exposed to frequent irrigation during summer months maintained high moisture levels during this period.

3) **Modeling soil moisture response to rainfall**: Using data collected from the monitoring network, a model was developed to predict soil moisture response to rainfall events at each of the instrumented green space sites, considering only rainfall depth and initial soil moisture as the model input parameters. This approach produced robust predictions for soil moisture increase at each site during wet-weather flow, with adjusted $R^2$ values exceeding 0.65, and effectively captured the complex
relationship between rainfall and soil moisture dynamics across a range of urban

green settings.

4) **Runoff generation and stormwater management:** The aforementioned model was
used to investigate how different monitored green spaces performed in terms of
runoff generation during rainfall events. It was found that while these spaces can
manage stormwater effectively for events up to 30-50mm, proactive soil drainage
systems are needed for green spaces to reduce stormwater runoff, and thus mitigate
flooding, during rainfall events over 50mm.

The research results associated with this particular study underscore the practical utility
of LoRaWAN technology in urban environmental monitoring. They also provide crucial insights
into the ability of different urban green spaces to mitigate stormwater runoff, especially for
larger storm depths. The results can be used to support city planners and environmental
managers tasked with optimizing the ecological benefits of green spaces and enhancing urban
resilience against climate change-induced challenges, including increased wet-weather flows.
5.2 Avenues for Future Research

For Chapter 2, while the findings underscored neighborhood income level as a significant factor influencing park usage by urban residents during the COVID-19 pandemic restrictions, these findings could be the result of a myriad of underlying factors. Prior research suggests that parks in lower-income areas often suffer from inadequate facilities, services, and maintenance. Beyond income level, attributes like sports facilities, water features, and aspects of the surrounding environment, including population and road density, proximity to an urban center, and accessibility via public transportation, are also known to impact park visitation. Future studies should delve deeper into these factors, providing more nuanced insights that could help enhance park services and design. Moreover, the observed changes in park visitation habits, varying by park type and socio-economic status, opens avenues for exploring how other demographic variables like age, gender, ethnicity, and transportation modes influence park visitation. Such comprehensive research would contribute significantly to understanding the evolving dynamics of urban park usage and inform strategies for creating more inclusive and responsive urban open spaces.

In Chapter 3, it was shown that the greenery surrounding a pixel of 30m×30m, influenced the LST of the pixel. The measure used to define “surrounding greenery” in this study was the Census Block Group (CBG) level runoff coefficient. However, given the diverse shapes and sizes of CBGs, this measure may only roughly approximate a pixel’s surroundings. Future research should explore the influence of surrounding greenery at different spatial scales and distances from a pixel. This could involve creating buffer zones of different radii around each pixel, calculating the runoff coefficient for these zones, and looking at the relationship between the runoff coefficient of these different sized zones and pixel LST. Such an approach would
enable the identification of an optimal influence zone or distance, whereby surrounding greenery (or hardscape) can influence the LST of a pixel, enhancing the accuracy of LST prediction models. It is noted, however, that considering the large number of pixel-level data generated for this study, this approach is likely to be computationally intensive and thus could require the employment of high-performance computing techniques. Nonetheless, adding this refining approach could yield more precise insights into the role of urban greening in mitigating urban heat island effects. The work presented in Chapter 3 could also be extended to other cities besides NYC, to explore how universal, or otherwise, the research findings are.

As for Chapter 4, future research can be conducted to both enhance the sensing system itself and to refine the modeling approach. One potential improvement for the sensing system is the integration of solar panels, as explored in previous studies, for supporting the energy needs to the system. This modification can potentially enable the system to become self-powered, significantly extending its operational lifespan in a single deployment. Other parameters that may influence the health and functionality of green spaces, such as solar radiation and CO₂ level, could be measured as well by adding corresponding sensors to the system.

Modeling approaches could also be expanded. While rainfall depth and initial soil moisture were established as key factors influencing soil moisture after a rainfall event, future models could incorporate elements like canopy interception, evaporation, and plant root water uptake. Moreover, estimating soil hydraulic properties, such as saturated hydraulic conductivity, from time-series soil moisture and meteorological data are also promising avenues to explore. Since these soil properties are incorporated in some established soil water retention models such as the HYDRUS, utilizing optimization techniques to tune the model parameters to match the
model predictions with the observed data could offer a more efficient alternative to traditional, resource-intensive measurement methods. If these model-based estimations are validated with in-situ measurements, they could greatly streamline soil health assessments and enhance the overall effectiveness of urban green space management.
5.3 Concluding Remarks

This thesis presents research that contributes to the knowledge advancement regarding the multifaceted challenges of urban sustainability in an ever-changing environment. Through a series of detailed studies, it delves into the complexities of urban park usage during the COVID-19 pandemic, the impact of urban greening on the urban heat island effect, and the use of LoRaWAN technology for monitoring urban green spaces. Each study contributes to a deeper understanding of how urban environments interact with social, ecological, and technological factors. The thesis not only highlights the dynamic nature of urban spaces during times of crisis but also underscores the potential of data-driven insights and technological advancements in enhancing urban resilience and sustainability. As cities continue to evolve, the findings and methodologies developed in this thesis illuminate the path for future research and practical applications. They serve as a valuable guide in confronting the challenges brought by climate change and urbanization, leading the way towards a more sustainable and resilient urban future.


120. Van Dyck, D. et al. Associations of neighborhood characteristics with active park use: an observational study in two cities in the USA and Belgium. Int. J. Health Geogr. 12, 26 (2013).


162. NYS DOT. Chapter 8 - Highway Drainage. in Highway Design Manual (New York State Department of Transportation, 2018).


### Appendix A: Supporting Information for Chapter 2

#### A.1 Summary of visit and visitor count by park type

Table A-1 Summary of visit and visitor count by park type

<table>
<thead>
<tr>
<th>Year</th>
<th>Park Type</th>
<th>NYC Local Visitor Count</th>
<th>US Visitor Count</th>
<th>All Visitor Count</th>
<th>All Visit Count</th>
<th>Local Visitor to US Visitor Fraction (%)</th>
<th>Local Visitor to All Visitor Fraction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019</td>
<td>Neighborhood Park</td>
<td>1,302,203</td>
<td>1,716,881</td>
<td>2,513,057</td>
<td>4,704,165</td>
<td>75.9</td>
<td>51.8</td>
</tr>
<tr>
<td></td>
<td>Community Park</td>
<td>1,009,907</td>
<td>1,291,056</td>
<td>1,708,520</td>
<td>3,285,743</td>
<td>78.2</td>
<td>59.1</td>
</tr>
<tr>
<td></td>
<td>Flagship Park</td>
<td>842,038</td>
<td>1,336,529</td>
<td>1,732,384</td>
<td>3,277,968</td>
<td>63.0</td>
<td>48.6</td>
</tr>
<tr>
<td></td>
<td>Jointly Operated Playground</td>
<td>763,718</td>
<td>875,623</td>
<td>1,227,608</td>
<td>2,924,808</td>
<td>87.2</td>
<td>62.2</td>
</tr>
<tr>
<td></td>
<td>Playground</td>
<td>663,992</td>
<td>805,454</td>
<td>1,153,740</td>
<td>2,357,716</td>
<td>82.4</td>
<td>57.6</td>
</tr>
<tr>
<td></td>
<td>Triangle/Plaza</td>
<td>391,581</td>
<td>574,661</td>
<td>831,113</td>
<td>1,272,079</td>
<td>68.1</td>
<td>47.1</td>
</tr>
<tr>
<td></td>
<td>Recreation Field/Courts</td>
<td>218,660</td>
<td>260,763</td>
<td>352,750</td>
<td>733,724</td>
<td>83.9</td>
<td>62.0</td>
</tr>
<tr>
<td></td>
<td>Nature Area</td>
<td>147,566</td>
<td>179,116</td>
<td>228,543</td>
<td>550,484</td>
<td>82.4</td>
<td>64.6</td>
</tr>
<tr>
<td>2020</td>
<td>Neighborhood Park</td>
<td>593,175</td>
<td>727,438</td>
<td>1,071,939</td>
<td>2,118,963</td>
<td>81.5</td>
<td>55.3</td>
</tr>
<tr>
<td></td>
<td>Community Park</td>
<td>554,014</td>
<td>643,666</td>
<td>888,053</td>
<td>1,786,854</td>
<td>86.1</td>
<td>62.4</td>
</tr>
<tr>
<td></td>
<td>Flagship Park</td>
<td>381,410</td>
<td>500,057</td>
<td>673,322</td>
<td>1,380,409</td>
<td>76.3</td>
<td>56.7</td>
</tr>
<tr>
<td></td>
<td>Jointly Operated Playground</td>
<td>374,947</td>
<td>440,377</td>
<td>638,408</td>
<td>1,411,035</td>
<td>85.1</td>
<td>58.7</td>
</tr>
<tr>
<td></td>
<td>Playground</td>
<td>365,820</td>
<td>455,150</td>
<td>699,710</td>
<td>1,375,563</td>
<td>80.4</td>
<td>52.3</td>
</tr>
<tr>
<td></td>
<td>Triangle/Plaza</td>
<td>132,099</td>
<td>178,209</td>
<td>248,510</td>
<td>398,149</td>
<td>74.1</td>
<td>53.2</td>
</tr>
<tr>
<td></td>
<td>Recreation Field/Courts</td>
<td>134,515</td>
<td>156,466</td>
<td>215,054</td>
<td>469,251</td>
<td>86.0</td>
<td>62.6</td>
</tr>
<tr>
<td></td>
<td>Nature Area</td>
<td>148,904</td>
<td>181,178</td>
<td>241,645</td>
<td>535,313</td>
<td>82.2</td>
<td>61.6</td>
</tr>
</tbody>
</table>
A.2 Temperature adjustment model for park visits

Table A.2 Relationship between the number of local NYC park visitors and temperature

<table>
<thead>
<tr>
<th>Park Type</th>
<th>Neighborhood Park</th>
<th>Community Park</th>
<th>Flagship Park</th>
<th>Jointly Operated Playground</th>
<th>Playground</th>
<th>Triangle/Plaza</th>
<th>Nature Area</th>
<th>Recreation Field/Courts</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson’s R</td>
<td>0.66</td>
<td>0.85</td>
<td>0.92</td>
<td>0.11</td>
<td>0.65</td>
<td>0.68</td>
<td>0.52</td>
<td>0.62</td>
<td>0.78</td>
</tr>
<tr>
<td>p value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.594</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.01</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Note: the sample sizes are all 26 (the number of months starting from Jan 2018 to Feb 2020)

To determine if there was a significant difference in temperature patterns between 2020 and 2019, a paired t-test was conducted for each month. The results, presented in Table S-3, revealed that temperatures in January, February, March, April, and November showed significant differences between the two years.

Table A.3 Paired t-test results for monthly temperatures in 2020 and 2019

<table>
<thead>
<tr>
<th>Month</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>31</td>
<td>28</td>
<td>31</td>
<td>30</td>
<td>31</td>
<td>30</td>
<td>31</td>
<td>31</td>
<td>30</td>
<td>31</td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td>T stat</td>
<td>2.77</td>
<td>2.24</td>
<td>3.66</td>
<td>-3.15</td>
<td>-1.15</td>
<td>1.54</td>
<td>0.39</td>
<td>1.65</td>
<td>-0.84</td>
<td>-0.93</td>
<td>5.17</td>
<td>0.65</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;0.01</td>
<td>0.03</td>
<td>&lt;0.001</td>
<td>&lt;0.01</td>
<td>0.26</td>
<td>0.13</td>
<td>0.70</td>
<td>0.11</td>
<td>0.41</td>
<td>0.36</td>
<td>&lt;0.001</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Therefore, to remove the effect of temperature on park visits, several models were developed using the ordinary least squares (OLS) method and the Gaussian Process method, as shown in Figure A-1.
The highest $R^2$ value of 0.72 was achieved by a 3rd-degree polynomial model ($f(t) = -24.4 \cdot t^3 - 325.2 \cdot t^2 + 15553.0 \cdot t + 318999.5$), which was then employed to adjust the park visits data.

$$f(t) = -24.4 \cdot t^3 - 325.2 \cdot t^2 + 15553.0 \cdot t + 318999.5 \quad (A-1)$$

Where $f(t)$ is the total number of park visits in a month, and $t$ is the average temperature in that month.

$$V_{base,adj} = V_{2019} \times (1 + \frac{f(t_{2020}) - f(t_{2019})}{f(t_{2020})}) \quad (A-2)$$

was used to adjust the number of monthly visits to every single park:
Where \( V_{2019} \) is the number of visits in 2019, \( t_{2019} \) and \( t_{2020} \) are the monthly average temperatures in 2019 and 2020, respectively, and \( V_{\text{base,adj}} \) is the assumed base number of visits after adjustment for temperature differences: this parameter can be interpreted as the number of visits to a park in a month of 2019 if the temperature in that month was the same as that in 2020.

The temperature adjustment model was developed for each of the four types of visits/visitors: 1) all visits, 2) all visitors, 3) US visitors, and 4) NYC local visitors, as described in the Data and Methods section. The model described in this section was developed using data for NYC local visitors; the modeling results by using the other three types of visits/visitors were all similar as the model provided here.
### A.3 The mean and change of travel distance by park type

Table A-4 Summary of the mean and change of travel distance by park type

<table>
<thead>
<tr>
<th>Park Type</th>
<th>Mean Travel Distance in 2019 (km)</th>
<th>Mean Travel Distance in 2020 (km)</th>
<th>Change of Travel Distance (km, 95% CI)</th>
<th>Change of Travel Distance (percentage, 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>5.9</td>
<td>5.1</td>
<td>-0.8 (-0.8, -0.8)</td>
<td>-13.2 (-13.4, -13.1)</td>
</tr>
<tr>
<td>Triangle/Plaza</td>
<td>7.1</td>
<td>6.2</td>
<td>-0.9 (-1.0, -0.9)</td>
<td>-13.1 (-13.8, -12.4)</td>
</tr>
<tr>
<td>Flagship Park</td>
<td>7.1</td>
<td>5.9</td>
<td>-1.2 (-1.3, -1.2)</td>
<td>-17.5 (-17.8, -17.1)</td>
</tr>
<tr>
<td>Recreation Field/Courts</td>
<td>6.0</td>
<td>4.9</td>
<td>-1.1 (-1.2, -1.1)</td>
<td>-18.6 (-19.3, -17.8)</td>
</tr>
<tr>
<td>Nature Area</td>
<td>6.0</td>
<td>6.0</td>
<td>-0.1 (-0.1, 0.0)</td>
<td>-1.2 (-2.0, -0.3)</td>
</tr>
<tr>
<td>Community Park</td>
<td>6.0</td>
<td>4.8</td>
<td>-1.2 (-1.2, -1.1)</td>
<td>-19.3 (-19.7, -19.0)</td>
</tr>
<tr>
<td>Neighborhood Park</td>
<td>5.9</td>
<td>5.0</td>
<td>-0.9 (-0.9, -0.9)</td>
<td>-15.0 (-15.3, -14.6)</td>
</tr>
<tr>
<td>Playground</td>
<td>5.0</td>
<td>5.0</td>
<td>0.1 (0.0, 0.1)</td>
<td>1.1 (0.6, 1.7)</td>
</tr>
<tr>
<td>Jointly Operated Playground</td>
<td>4.5</td>
<td>4.4</td>
<td>-0.1 (-0.1, -0.1)</td>
<td>-1.9 (-2.5, -1.3)</td>
</tr>
</tbody>
</table>
### A.4 The mean and change of travel distance by park type and by income level of visitor CBGs

#### Table A-5 Summary of the mean and change of travel distance by park type and by income level of visitor CBGs

<table>
<thead>
<tr>
<th>Park Type</th>
<th>Mean Travel Distance – 2019, 2020 (km)</th>
<th>Change of Travel Distance and 95% CI (km)</th>
<th>Change of Travel Distance and 95% CI (percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>lower, middle, upper</td>
<td>lower, middle, upper</td>
<td>lower, middle, upper</td>
</tr>
<tr>
<td>Overall</td>
<td>5.3, 4.7</td>
<td>-0.6, -0.9</td>
<td>-10.7, -13.9, -15.8</td>
</tr>
<tr>
<td>Community Park</td>
<td>6.5, 5.6</td>
<td>6.1, 4.5</td>
<td>11.7, 20.7, 25.7</td>
</tr>
<tr>
<td>Flagship Park</td>
<td>6.1, 5.5</td>
<td>7.5, 5.8</td>
<td>9.0, 17.8, 23.6</td>
</tr>
<tr>
<td>Jointly Operated Playground</td>
<td>4.4, 4.3</td>
<td>4.5, 4.3</td>
<td>-2.2, 0.2, -3.7</td>
</tr>
<tr>
<td>Nature Area</td>
<td>7.2, 7.4</td>
<td>5.7, 5.1</td>
<td>3.0, 9.4, -9.8</td>
</tr>
<tr>
<td>Neighborhood Park</td>
<td>5.3, 4.6</td>
<td>5.6, 4.9</td>
<td>-13.6, -18.6, -13.3</td>
</tr>
<tr>
<td>Playground</td>
<td>4.0, 3.8</td>
<td>6.2, 6.5</td>
<td>-4.5, 6.2, 4.7</td>
</tr>
<tr>
<td>Recreation Field/Courts</td>
<td>5.5, 4.2</td>
<td>6.0, 5.4</td>
<td>-23.8, -18.9, -9.8</td>
</tr>
<tr>
<td>Triangle/Plaza</td>
<td>7.3, 6.2</td>
<td>5.1, 4.2</td>
<td>-15.8, -9.0, -18.2</td>
</tr>
</tbody>
</table>

Note: The values in parentheses indicate the change in travel distance and its 95% confidence interval (CI).
### A.5 The mean and change of travel distance by park type and by income level of park CBGs

#### Table A-6 Summary of the mean and change of travel distance by park type and by income level of park CBGs

<table>
<thead>
<tr>
<th>Park Type</th>
<th>Mean Travel Distance – 2019, 2020 (km)</th>
<th>Change of Travel Distance and 95% CI (km)</th>
<th>Change of Travel Distance and 95% CI (percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>lower</td>
<td>middle</td>
<td>upper</td>
</tr>
<tr>
<td>Overall</td>
<td>5.7, 4.9</td>
<td>5.6, 4.7</td>
<td>6.3, 5.6</td>
</tr>
<tr>
<td>Community Park</td>
<td>6.5, 4.9</td>
<td>5.3, 4.5</td>
<td>5.9, 5.0</td>
</tr>
<tr>
<td>Flagship Park</td>
<td>7.1, 5.9</td>
<td>8.3, 5.7</td>
<td>6.5, 5.9</td>
</tr>
<tr>
<td>Jointly Operated Playground</td>
<td>4.4, 4.7</td>
<td>4.7, 3.9</td>
<td>4.3, 4.4</td>
</tr>
<tr>
<td>Nature Area</td>
<td>5.7, 6.1</td>
<td>5.9, 5.5</td>
<td>6.2, 6.2</td>
</tr>
<tr>
<td>Neighborhood Park</td>
<td>5.9, 4.8</td>
<td>4.9, 4.5</td>
<td>6.5, 5.8</td>
</tr>
<tr>
<td>Playground</td>
<td>4.4, 4.9</td>
<td>5.2, 4.9</td>
<td>6.1, 5.6</td>
</tr>
<tr>
<td>Recreation Field/Courts</td>
<td>6.0, 4.7</td>
<td>6.0, 5.1</td>
<td>6.2, 5.3</td>
</tr>
<tr>
<td>Triangle/Plaza</td>
<td>5.4, 4.3</td>
<td>5.1, 4.5</td>
<td>7.8, 7.1</td>
</tr>
</tbody>
</table>
A.6 Auxiliary information about park characteristics

Figure A.2 (a) Income level of park CBGs by park type; (b) park area by park type. The letters inside the boxes are the Tukey HSD multi-group comparison results for each topic, any common letter shared by two park types indicates that the two park types were found to belong to the same group.
## A.7 Park classification standard

This park classification standard is directly quoted from the NYC Department of Parks and Recreation, which is available at [this link](#).

### Table A-7 Park classification standard by the NYC Department of Parks and Recreation (DPR)

<table>
<thead>
<tr>
<th>Park Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community Park</td>
<td>Parks with a combination of active and passive recreational facilities or specialized amenities that service more than one community district. The largest of these parks have a large natural area component. Community Parks are typically between 5 and 250 acres.</td>
</tr>
<tr>
<td>Flagship Park</td>
<td>These parks have a variety of active recreational facilities and include a large natural or landscaped component. Flagship Parks are some of the largest parks in the City and attract users from throughout the metropolitan region. They include, but are not limited to those joint interest areas that traverse multiple community districts.</td>
</tr>
<tr>
<td>Jointly Operated Playground</td>
<td>Playgrounds adjacent to public schools jointly operated by NYC DPR and the Department of Education</td>
</tr>
<tr>
<td>Nature Area</td>
<td>Vacant/unimproved area, including islands, which are not associated with other parks and contain natural features including forests, marshland, meadows, etc.</td>
</tr>
<tr>
<td>Neighborhood Park</td>
<td>Parks that are intended to serve the direct neighborhood in which they are located. Neighborhood Parks are typically up to 50 acres and may include passive/active recreational areas.</td>
</tr>
<tr>
<td>Playground</td>
<td>Standalone facilities under DPR jurisdiction and management consisting of playground equipment along with perhaps hard surface or turf sports areas. Typically under 5 acres and more than 50% of the total site.</td>
</tr>
<tr>
<td>Recreational Field/Courts</td>
<td>Sites consisting solely of hard surface/turf sports areas that are operated by DPR.</td>
</tr>
<tr>
<td>Triangle/Plaza</td>
<td>A landscaped or paved area usually in conjunction with the arterial or local street system. These sites are primarily developed for passive recreation use or to provide an open space amenity within the neighborhood. These sites are nonlinear and may be developed to contain grass, trees, shrubs, cobblestone, fences, monuments, plaques, flagpoles, benches, game or picnic tables and drinking fountains, but they have no active recreational equipment. Smaller sites may be operated as Greenstreets, but are under the jurisdiction of DPR. These sites range in size, but are typically under 1 - acre.</td>
</tr>
</tbody>
</table>
A.8 Administrative orders regarding park/school closure and reopening in NYC

The timeline of these administrative orders was compiled from the government website and their social media account, where these orders were announced.

![Timeline of the closure and reopening of NYC parks](image1)

Figure A-3 Timeline of the closure and reopening of NYC parks

![Timeline of the closure and reopening of NYC schools](image2)

Figure A-4 Timeline of the closure and reopening of NYC schools
A.9 Assessment about normalizing the visits for the change in SafeGraph’s device panel

Figure A-5 Example figures of raw visits data vs. normalized visits data by total visitors. (a) Flagship Park (b) Nature Area

Table A-8 Pearson’s correlation coefficient between raw visits and normalized visits by total visitors for each park type

<table>
<thead>
<tr>
<th>Park Type</th>
<th>Community Park</th>
<th>Flagship Park</th>
<th>Jointly Operated Playground</th>
<th>Nature Area</th>
<th>Neighborhood Park</th>
<th>Playground</th>
<th>Recreation Field/Courts</th>
<th>Triangle/Plaza</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson’s R</td>
<td>0.964</td>
<td>0.982</td>
<td>0.972</td>
<td>0.763</td>
<td>0.974</td>
<td>0.919</td>
<td>0.923</td>
<td>0.990</td>
<td>0.980</td>
</tr>
</tbody>
</table>
Appendix B: Supporting Information for Chapter 4

B.1 LoRaWAN signal survey in Upper Manhattan

Figure B-1 Map of the LoRaWAN signal survey in Upper Manhattan