EXAMINING HOW WEATHER AND OUTDOOR THERMAL COMFORT VARIABILITY AFFECTS CYCLING ACTIVITY USING BIG DATA AND MICROCLIMATE SIMULATIONS IN NEW YORK CITY

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Anna Makido

August 2019

ABSTRACT

Urbanization is rapidly occurring and with this growth in urban populations, it is crucial to plan cities carefully to create a livable and healthy living environment. To better understand human-environment relationships and increase eco-friendly transportation modes, previous studies have discussed the effects of weather variability on urban transportation modes. This study adds to the existing literature by running multiple regressive models to examine the weather-cycling relationship using a variety of temporal scales. Additionally, microclimate simulations were conducted to calculate location specific Universal Thermal Comfort Index (UTCI) values in New York City. Findings indicate that cyclists are more vulnerable to weather variability during the Spring and Fall seasons. The regression results for UTCI suggest that outdoor thermal comfort can be used as a predictor for cycling activity. Furthermore, the simulated location specific UTCI values displayed a stronger effect on bike usage. These findings highlight the importance of conducting microclimate simulations.

BIOGRAPHICAL SKETCH

Anna Makido was born in Japan and mostly raised in Michigan but had the opportunity to live in a broad range of places until eventually moving to Oregon in high school. Although Anna hoped to study Psychology in college, her plans changed when she saw a documentary about the effects of climate change. Devasted by a scene portraying a polar bear stranded on a small piece of ice, she immediately decided to change her career goals and became interested in the environment. Upon graduating high school, Anna continued her education and graduated with a Bachelor of Science Degree in Environmental Geosciences from Purdue University. Through the courses at Purdue University, Anna gained interest in urban environments, especially in human-environment interactions. Anna studied abroad during her third year at Osaka University, where she deepened her knowledge in architectural planning, urban design, energy and environmental issues, and environmental behavior. In 2017, Anna moved to Ithaca, New York to join the Master of Science in Regional Science program at Cornell University to pursue her interests. Anna's research interests include sustainable urban development, green buildings, and environmental psychology. Anna's long-term goal is to promote the implementation of green and livable cities around the world.

ACKNOWLEDGMENTS

This thesis would not have been possible without the support of many individuals.

I would first like to thank my committee members, Kieran Donaghy and Timur Dogan for their continuous guidance not only for my thesis, but during my two years at Cornell. Kieran has been supportive from the first time I visited Cornell and has been instrumental in shaping my academic and professional development. I am thankful for all the wise advice that he has given me, and for always pushing me to pursue my interests. I would also like to thank Timur for his ideas and enthusiasm that helped me make progress in this thesis. I could not have completed this thesis without the advice that he has given me throughout this process.

I would like to thank Patrick Kastner for his tremendous help in the simulations, and for always being patient in answering all my questions. I would also like to thank my roommate, Edith Zheng for always being so kind in helping me with (and teaching me) statistics late at night.

I would like to thank my parents for their continuous support. I thank my mother for being the best role model, and for giving me so much knowledge about the world. Thank you for the endless love, believing in me, and always being there.

Finally, I would like to thank my lovely friends for all of the emotional support during my two years at Cornell. Thank you for always making me laugh and giving me a home away from home. I will forever cherish the wonderful memories that I made with them.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

In 1800, only 3 percent of the world's population lived in urban areas; by 1900 this number increased to 14 percent; and in 2008 the world's population was evenly split between urban and rural areas (Population Reference Bureau). Urbanization is rapidly occurring, and it is expected that by 2050, 70 percent of the world's population will live in urban areas. With this growth in urban population, it is crucial to plan cities carefully to create a livable and healthy living environment. In response to this growing attention in creating sustainable cities, numerous cities have announced sustainable development goals. New York City is one of them and has created the OneNYC goal of reducing greenhouse gas (GHG) emission 80 percent from 2005 levels by 2050 as a means to combat climate change (The City of New York, Mayor's Office of Sustainability). As of 2015, NYC emissions have dropped 14.8 percent since 2005, which means there is still a need for a 65.2 percent reduction by 2050. Although New York City buildings were responsible for a large portion of GHG emissions in 2015 (67 percent), it is important to note that the transportation sector accounted for 30 percent of GHG emissions. In the updated OneNYC 2050 strategy, some of the goals include "achieve carbon neutrality and 100 percent clean electricity" under the A Livable Climate initiative and "reduce congestion and emissions" under the Efficient Mobility initiative (City of New York). Increasing alternative methods of transportation such as cycling, can help achieve these goals.

In comparison with other travel modes, including cars, trains, and buses, cycling is more likely to be affected by weather as cyclists are exposed to the outdoor environment (Liu et al., 2015a; Sabir, 2011). Although the relationship between urban transportation and weather has been gaining attention (Guo et al., 2007; Kalkstein et al., 2009; Liu et al., 2015a; Miranda-Moreno and Nosal, 2011; Singhal et al., 2014; Thomas et al., 2013; Zhao et al., 2018), it is important to further strengthen the understanding of the relationship between weather and cycling. Many of these previous studies used field survey data as opposed to big data. Now with the increasing availability of open data, it is possible to obtain temporally and spatially dense data. Very few studies have explored cycling big data to compare weather-cycling relationships (Zhao et al., 2018). Furthermore, numerous studies have compared weather variables with cycling, but few have explored the relationship between outdoor thermal comfort and cycling activities. Specifically, the Universal Thermal Comfort Index (UTCI), a measure of outdoor thermal comfort first introduced in 2012 which has recently been gaining recognition as an accurate method to measure outdoor thermal comfort. As such it is particularly important to have an in-depth understanding of outdoor thermal comfort and cycling relationships so that the potential reduction in cycling due to uncomfortable outdoor environments can be mitigated.

This study aims to fill these research gaps and add to the existing literature by adding a spatial component to the statistical analysis using cycling and weather data. Cycling and weather data have been obtained through Citi Bike and Weather Underground, respectively, and multiple regressive models were estimated to gain an understanding of the relationship. Several temporal scales were used in examining the relationship between weather and cycling, as well as UTCI and cycling. Finally, a microclimate simulation for Manhattan was conducted to calculate the UTCI values for specific Citi Bike stations to explore the spatial component of the relationship between weather, UTCI, and cycling. The results of the study will help inform decision makers of the most important environmental variables affecting cycling and suggest ways to alter the built environment to create a more comfortable environment for outdoor activities. This framework will help planners and urban designers in prioritizing what to include to create an attractive street to increase outdoor activities.

CHAPTER 2

LITERATURE REVIEW

There have been few studies directly quantifying the relationship between outdoor activities and climate conditions. Moreover, even fewer of these studies examine the relationship between outdoor thermal comfort values and outdoor activities. Among the existing studies exploring weather and transportation, there have been common findings. Precipitation generally always has a negative effect on ridership across all transit modes (Guo et al., 2007; Kalkstein et al., 2009; Zhao et al., 2018). Since the effect of precipitation on cycling has already been explored thoroughly in previous studies with consistent results (Bergstrom and Magnusson, 2013; Winter et al., 2017; Zhao et al.,), this study does not include precipitation as a variable in the regression models.

Understanding the behavioral impacts of weather variability on travel behavior is critical for planners and policymakers to incorporate when creating design guidelines and infrastructure management. Previous studies have examined the impacts of travel purposes on the relationship between weather variability and travel behavior. These studies have found that commuters are less sensitive than non-commuters to weather variability (Cools et al., 2010; Liu et al., 2015). Furthermore, studies have shown that adverse weather strongly influences cycling (Richardson, 2000; Bergstrom and Magnusson, 2003; Winter et al., 2007). This is reasonable as cyclists are less protected from the outdoor environment compared to other modes of transportation. A study conducted in Seattle, Washington concluded that cycling is largely self-dependent at finer temporal scales (Zhao et al., 2018). Another study introduced the concept of a 9-term average residual in order to mitigate the temporal variations of the variables (Kalkstein et al., 2009). This 9-term average residual is widely used by other studies examining weather and transit relationships (Singhal et al., 2014; Zhao et al., 2018).

Although previous studies have quantitatively examined the relationship between weather and travel patterns, most studies have focused on public transportation mode choices instead of transportation modes that are more prone to weather variability such as walking or cycling largely due to the lack of temporally and spatially fine data. Additionally, most existing studies examining UTCI have focused on the validation of the thermal comfort index as opposed to the application of it (Pantavou et al., 2013; Lai et al., 2014; Fang et al., 2018). This study aims to fill the research gaps by quantitatively provide an in-depth understanding of the relationship between outdoor thermal comfort and cycling activities.

CHAPTER 3

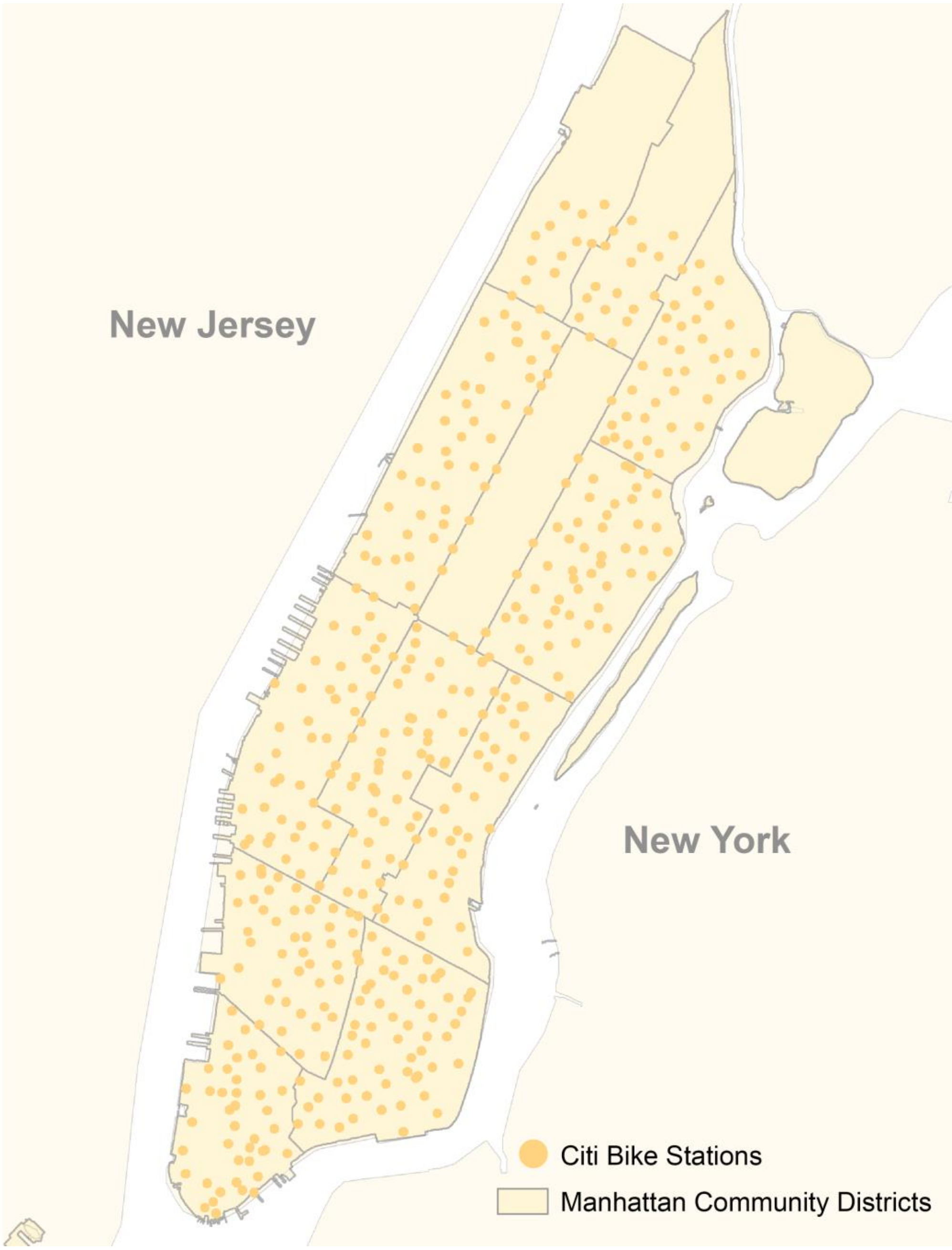
STUDY CONTEXT AND DATA SOURCES

3.1 Study Area

Manhattan, New York was chosen as the study area. Manhattan is the most densely populated Borough within New York City, and is also a popular destination for tourists. The island is enclosed by the Hudson, East, and Harlem rivers. The population of Manhattan is estimated to be around 1.6 million, with nearly 62.8 million tourists in 2017 (NYC Future). New York City is situated in the warm humid subtropical climate zone, where winter temperatures generally stay a few degrees above freezing, and summer conditions exist around late May to late September (National Climatic Data Center). New York City is generally not very hilly, making it an ideal location for cyclists.

The large number of visitors and residents in Manhattan allow for an interesting mix of data through Citi Bike. The Department of Transportation (DOT) in NYC has released Safer Cycling, a comprehensive study of NYC's bicycle network, which reports that numbers of regular bicyclists have increased, and cycling has grown dramatically safer. 24 percent of New Yorkers, nearly 1.6 million people report to have cycled at least once a year (City of New York), and on a typical day there are about 490,000 cycling trips made in NYC (U.S. Census Bureau, 2017). In an effort to accommodate the growing number of cyclists, NYC DOT continues to expand and enhance on-street bike networks as well as dedicated bike lanes.

Figure 1: Context map of Manhattan showing Community Districts and bike stations



3.2 Weather Data

The 2018 weather data were obtained through the Weather Underground website. This data source has been used extensively by previous studies, and research has indicated that public domain weather sources offer weather data that are just as reliable as an onsite weather station (Singhal et al., 2014; Wolff et al., 2011). Since the weather data obtained through Weather Underground would be used as an input for the microclimate simulations, as well as serve as an overall weather for NYC, a station that was not in another building's shadow was needed. In order to ensure that the weather station was placed at an unobstructed area, a station in Jackson Heights, New York [Station ID: KNYJACKS2] was used to obtain data. The hardware used for this station is Davis Vantage Pro, and the software used is meteobridge. The weather factors acquired from this data source include air temperature, relative humidity, wind direction, wind speed, solar radiation, and precipitation. The raw data obtained through Weather Underground had a temporal resolution of 15 minutes, so the data were first aggregated into hourly values.

3.3 Cycling Data

Cycling data were obtained through the Citi Bike NYC Trip Histories Data website. Citi Bike launched in 2013 and is now the largest bike share system in the nation. Stations and thousands of bicycles are placed across Manhattan, Brooklyn, Queens, and Jersey City, and are available for use anytime. Citi Bike offers both day passes and annual memberships, attracting both tourists and residents in New York. The information included in the Citi Bike Data include trip duration, start and stop time and date, start and end station coordinates, and user type. The user type data include information such as age, gender, and whether the biker user is a one-time use customer or a subscriber. The raw data included all trip logs, so the data were first aggregated into hourly counts using R software.

3.4 Simulation Software for Microclimate Modeling

For the microclimate simulations, a 3D model of the buildings in NYC was downloaded from the NYC Planning website. The bike station points were obtained through the latitude and longitude coordinates provided by the Citi Bike data. To ensure that the station points had the same projection as the 3D model, the station points were assigned a projection using ArcGIS. The shape file that resulted from this process was then put back into Rhino, and the 3D model was aligned with the station location points. In an effort to speed up the simulation process, the 3D model was split into the 12 community districts within Manhattan. Table 1 is a list of the community districts in Manhattan.

Table 1: List of neighborhoods in each Community District

Community District	**Neighborhood Names**
MN CD 01	Battery Park City, Civic Center, Ellis Island, Governors Island, Liberty Island, South Street Seaport, Tribeca, Wall
MN CD 02	Greenwich Village, Hudson Square, Little Italy, NoHo, SoHo, South Village, West Village
MN CD 03	Chinatown, East Village, Lower East Side, NoHo, Two Bridges
MN CD 04	Chelsea, Clinton, Hudson Yards
MN CD 05	Flatiron, Gramercy Park, Herald Square, Midtown, Midtown South, Murray Hill, Times Square, Union Square
MN CD 06	Beekman Place, Gramercy Park, Murray Hill, Peter Cooper Village, Stuyvesant Town, Sutton Place, Tudor City, Turtle Bay
MN CD 07	Lincoln Square, Manhattan Valley, Upper West Side
MN CD 08	Carnegie Hill, Lenox Hill, Roosevelt Island, Upper East Side, Yorkville
MN CD 09	Hamilton Heights, Manhattanville, Morningside Heights, West Harlem
MN CD 10	Central Harlem
MN CD 11	East Harlem, Harlem, Randall's Island Park, Wards Island Park
MN CD 12	Inwood, Washington Heights
MN Parks	Central Park

CHAPTER 4

METHODOOGY: STATISTICAL ANALYSIS

4.1 Independent Variables

Weather variables have an inherent time series variation resulting from diurnal cycles. To control for this temporal variation, the residual weather equation was used to calculate the residuals for air temperature, relative humidity, wind speed, and solar radiation (Zhao et al., 2018). The residuals allow for a comparison of data across all days of the week (Kalkstein et al., 2009).

The formula for the residual weather variables is:

$$\Delta W_t = \frac{W_t - \overline{W}_t^{MA\ \pm 4}}{\overline{W}_t^{MA\ \pm 4}}$$

where

$$\overline{W}_t^{MA\ \pm 4} = \frac{\sum_{\tau=-4}^{4} W_{t+7\tau}}{9}$$

ΔW_t is the residual weather variables of air temperature, relative humidity, wind speed and solar radiation. W_t is the observed weather conditions at a particular day or hour t. $\overline{W}_t^{MA\ \pm 4}$ is the 9-term moving average for day or hour t. The index τ represents weeks, ranging from -4 to 4. Therefore $W_{t+7\tau}$ ranges from 28 days before to 28 days after the day or hour in question. The weather variable value for the day or hour in question will only be compared to a similar time of year. This range of 4 weeks before and after was chosen as it is not too long so that it would include days that are substantially different from the day in question such as changes in seasons and academic year (Kalkstein et al., 2009).

4.2 Dependent Variable

Similar to the residuals weather variables, residual variables for cycling counts were calculated using the following formula:

$$\Delta C_t = \frac{C_t - \bar{C}_t^{MA\,\pm 4}}{\bar{C}_t^{MA\,\pm 4}}$$

where

$$\bar{C}_t^{MA\,\pm 4} = \frac{\sum_{\tau=-4}^{4} C_{t+7\tau}}{9}$$

ΔC_t is the residual cycling counts at either the daily or hourly level. C_t is the observed weather conditions at a particular day or hour t. $\bar{C}_t^{MA\,\pm 4}$ is the 9-term moving average for day or hour t. The index τ represents weeks, ranging from -4 to 4.

By taking the residuals of the cycling counts, the time series variation of cycling by day of time, time of day, and other non-weather effects can be controlled. For example, mornings and evenings will always have higher cycling counts on the weekdays and the afternoons will have higher cycling counts on the weekends. The residuals will mitigate the effect of these non-weather related temporal variations.

4.3 Multivariate Linear Regression

The multivariate regression model used for the statistical analysis is shown below:

$$\Delta C_t = b_0 + b_1 \Delta T_t + b_2 \Delta WS_t + b_3 \Delta Sol_t + b_4 \Delta RH_t$$

where ΔC_t is the residual cycling counts at the daily or hourly level, ΔT_t is the daily or hourly residuals for air temperature, ΔWS_t is the daily or hourly residuals for wind speed, ΔSol_t is the daily or hourly residuals for solar radiation, and ΔRH_t is the daily or hourly residuals for relative humidity. A summary of the models is shown in Table 3. Although some previous studies have included a lag variable into the regression model (Zhao et al., 2018), the residuals calculated with the 9-term moving average already account for temporal variations, so it is not included in this regression model.

Different temporal scales were used to estimate the regressions models, including daily versus hourly values, weekdays versus weekends, and the four seasons. The seasons were categorized using the bins shown in Table 2.

Table 2: List of months per season

Season	**Months**
Winter	December, January, February
Spring	March, April, May
Summer	June, July, August
Fall	September, October, November

Table 3: Summary of non-spatial models

	Dependent Variable (cycling counts)	**Independent Variable**
Model 1	Hourly / Daily values for 2018 for all users	Weather Variables
Model 2	Hourly/ Daily values for 2018 for 'customers'	Weather Variables
Model 3	Hourly / Daily values for weekends	Weather Variables
Model 4	Hourly/ Daily values for weekdays	Weather Variables
Model 5	Hourly / Daily values for 4 seasons	Weather Variables
Model 6	Hourly / Daily values for 2018 for all users	UTCI Values

CHAPTER 5

METHODOLOGY: SPATIAL ANALYSIS

5.1 Microclimate Modeling

The input variables for UTCI calculations were simulated using Grasshopper components in Rhino. The diffuse and direct radiation which were needed for Mean Radiant Temperature (MRT) calculations, were simulated using the DIVA 4.0 Radiation Map component. The bike station locations in the form of point data were used as the sensor locations, and the radiation values were calculated at these locations. The buildings in the 3D model were also used as inputs for the radiation calculations to account for shadows that influence direct radiation, and scattered radiation from building walls which influence diffuse radiation. The wind velocities were calculated using Eddy, a Grasshopper plugin tool. The outputs for both radiation values and wind velocities were obtained for all hours in 2018. These simulated values were then used as inputs for the UTCI component. The UTCI component requires air temperature, relative humidity, wind speed, diffuse radiation, direct radiation, sun elevation angle, and average surface temperature context as inputs. The diffuse and direct radiation are used to calculate MRT, which is the average temperature of an imaginary enclosure which exchanges thermal radiation with the human body. The MRT is important in determining thermal comfort, as the human body is constantly exchanging thermal radiation with the surroundings. The air temperature, relative humidity, and sun elevation angle values from the weather station data obtained from

Weather Underground were used as input values. The default value was used for the average street width and height, body solar absorption, and ground reflectance. The output of this grasshopper workflow was hourly UTCI values in 2018 at each bike station. Simulations were conducted for each community district in Manhattan and were combined after all simulation results were obtained.

5.2 Universal Thermal Climate Index

UTCI was created as a means of creating a thermal index that could be used globally and accepted internationally. It is defined as “the air temperature of a reference condition causing the same model response as actual conditions” (Blazejczyk et al., 2013). The UTCI values would differ depending on the location of the bike stations because factors such as wind speed and solar radiation would vary. For example, if a station is located in between skyscrapers, that station would receive very little solar radiation and high wind speeds. These conditions would significantly decrease the UTCI value at that location, which may make it seem too cold and undesirable for Citi Bike customers. Comfortable UTCI values would result in more Citi Bike users, because the nice outdoor conditions may convince people to bike instead of taking an Uber or public transportation.

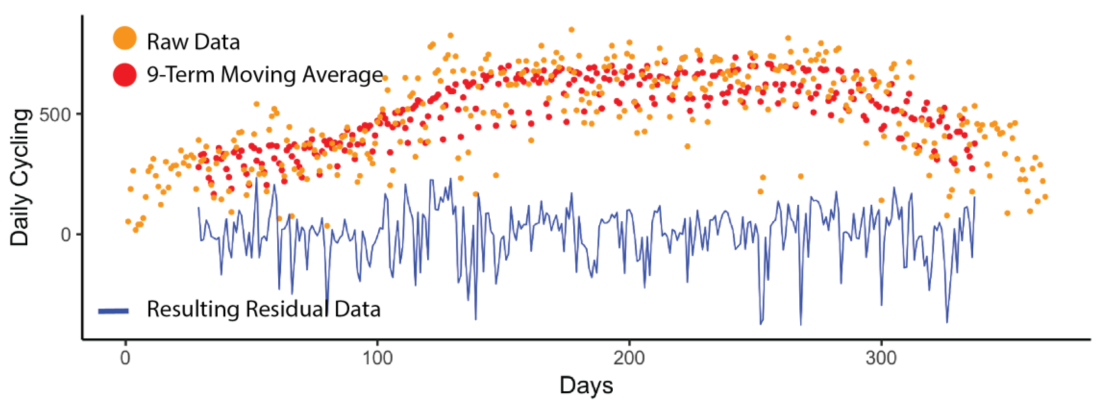
CHAPTER 6

RESULTS: STATISTICAL ANALYSIS

6.1 Descriptive Statistics

As seen in Figure 2 below, cycling counts gradually increase from January to July where it plateaus, and then decreases again moving towards December. The orange dots show raw cycling count data, the red dots shows the 9-term moving average, and the blue line illustrates the residual cycling values for the year 2018. The temporal pattern described above is no longer apparent when looking at the residual cycling values. This indicates that the temporal pattern was successfully reduced by taking the residual values, and these residual values are a better representation of weather-related effects.

Figure 2: Graph of cycling counts raw data, 9-term moving average, and residuals



Examining the cycling counts at a finer temporal resolution, we can observe a two-peak pattern for weekdays (Figure 3) and a one peak pattern for weekends (Figure 4). The two-peak pattern apparent on weekdays can be explained by morning and evening commute times. Based on this pattern, a large portion of Citi Bike users on weekdays seem to be commuters, and Citi Bike users on the weekends are cycling more for leisure. This difference in type of customers may affect how the bike users are influenced by the weather, and so a regression analysis is also conducted separately for weekdays and weekends. The varying nature of cycling counts depending on the day of the week and seasons further emphasizes the importance of exploring weather-cycling relationships at various temporal levels.

Figure 3: Two-peak pattern for hourly bike trips data on April 26th

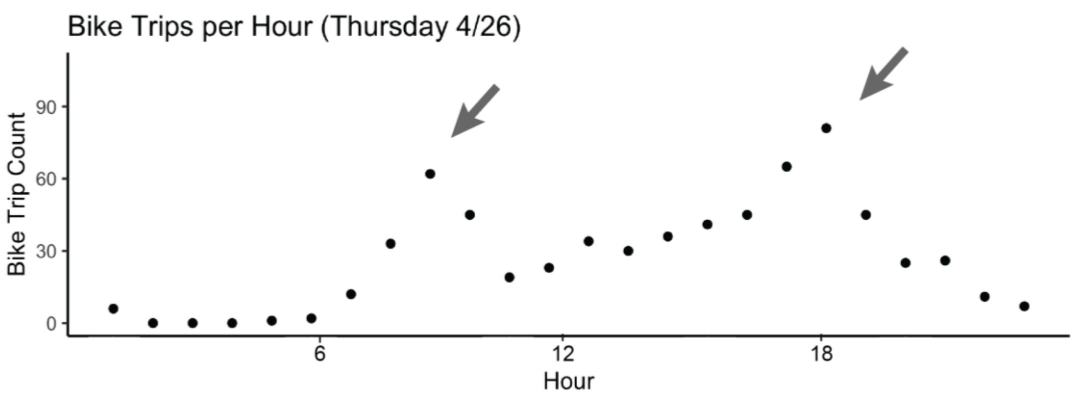
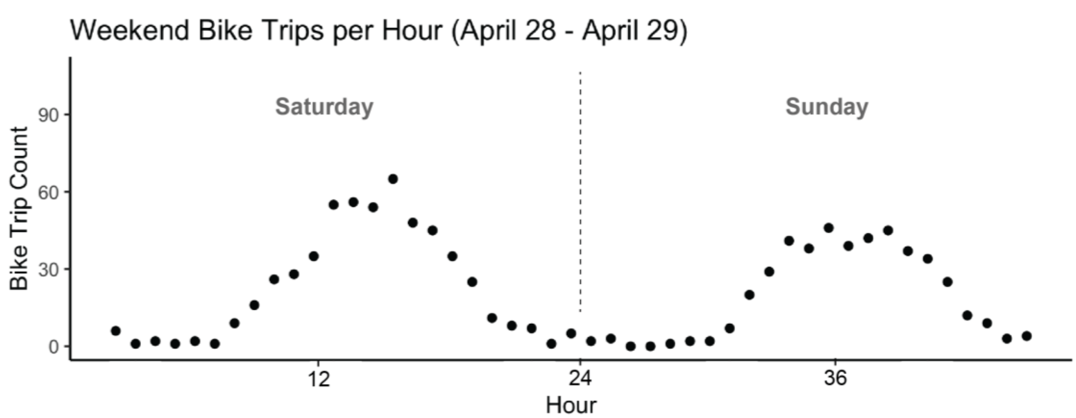


Figure 4: One-peak pattern for hourly bike trips data from April 28th to April 29th



6.2 Co-linearity Diagnostics

Before running the regression models, a multicollinearity test was conducted using the Variation Inflation Factor (VIF). The VIF measures how much the variance of an estimated regression coefficient increases if the predictors are correlated. If none of the factors are correlated, the VIFs would all be 1. A VIF between 5 and 10 indicates a high correlation. All variables were well below the collinearity diagnostics thresholds, and therefore could be included into the regression models. The result of the VIF test is shown in Table 4.

Table 4: VIF results

	ΔT_t	ΔWS_t	ΔSol_t	ΔRH_t
VIF	1.013201	1.009203	1.016188	1.029121

6.3 Regression Model Results

Daily and hourly cycling models are developed for various time periods throughout the year. Models were created to test the effects of the day of the week as well as seasonal changes. All estimated models are statistically significant as demonstrated by the F statistics and the P values. The R squared value slightly increased from 0.3852 to 0.4105 when only considering ‘customer’ users for the daily model. This suggests that ‘customer’ users are more likely to change their transportation modes based on the weather. A ‘customer’ is defined as someone who used either a 24- hour pass or a 3-day pass, as opposed to a ‘subscriber’ who has an annual membership. Perhaps a 'subscriber' is less prone to weather variability because they have the annual membership in order to use Citi Bike for commuting purposes. Commuters are less likely to have the luxury to change transportation modes or shift their cycling time since they do not have a flexible time schedule. Interestingly, the R squared value significantly decreases from 0.06547 to 0.04532 when only considering 'customer' users for the hourly model. This change in R squared value is larger than the difference in R squared values for the daily model. The results for the hourly model 1 and model 2 regression estimation suggest that 'customers' are less likely to be affected by weather variability. This illustrates the importance of conducting regression models using various temporal scales.

In terms of the effects of each weather variable, air temperature, wind speed, and relative humidity were significant. Air temperature had a positive relationship with cycling activity, while relative humidity and wind speed had a negative relationship with cycling activity. Moreover, relative humidity has the strongest effect on cycling for both hourly and daily models, as shown by the beta coefficient values.

Comparing the weekend and weekday models, the weather variables in the weekend model explain higher variation than in the weekday model for hourly values. Cycling activity on weekdays tend to be more severely affected by temperature, as can be seen from the beta coefficient values. Similar to the results from Model 1 and Model 2, the weekday model is less likely to be affected by weather than weekend cycling at the hourly level, but the opposite is true at the daily level. The daily model regression results suggest that cyclists are more likely to be influenced by weather variables on weekdays than the weekend.

Table 5: Model 1 and Model 2 regression beta coefficient results

	Hourly (All users)	***Hourly (Customers)***	***Daily (All users)***	***Daily (Customers)***
Intercept	-0.0046	-0.0243	0.0005	-0.7836*
Residual Ta	0.1056*	0.3645*	0.2768*	0.1410*
Residual WS	-0.1247*	-0.1302*	-0.1780*	-0.0654*
Residual Solar	-0.0092	-0.0112	-0.0298	0.0144
Residual RH	-0.4903*	-0.6588*	-0.5278*	-0.1502*
Adjusted R squared	0.0655	0.0453	0.3852	0.4105
Model F	123.1	75.36	47.84	53.05

* $p < .05$

Comparing the weekend and weekday models, the weather variables in the weekend model explain higher variation than in the weekday model for hourly values. Cycling activity on weekdays tend to be more severely affected by temperature, as can be seen from the beta coefficient values. Similar to the results from Model 1 and Model 2, the weekday model is less likely to be affected by weather than weekend cycling at the hourly level, but the opposite is true at the daily level. The daily model regression results suggest that cyclists are more likely to be influenced by weather variables on weekdays than the weekend.

Table 6: Model 3 and Model 4 daily regression beta coefficient results

	Weekday (Daily)	***Weekend (Daily)***
Intercept	0.00364	-0.00731
Residual Ta	0.32154*	0.16187*
Residual WS	-0.22750*	-0.08720
Residual Solar	-0.10463	0.07007
Residual RH	-0.56865*	-0.48940*
Adjusted R squared	0.45620	0.25640
Model F	45.26	8.498

* $p < .05$

Table 7: Model 3 and Model 4 hourly regression beta coefficient results

	Weekday (Hourly)	***Weekend (Hourly)***
Intercept	0.00035	-0.01749
Residual Ta	0.11293*	0.08667*
Residual WS	-0.12516*	-0.12467*
Residual Solar	-0.01164	-0.00290
Residual RH	-0.48545*	-0.49637*
Adjusted R squared	0.06173	0.07615
Model F	82.42	42.65

* $p < .05$

For the seasonal model, Fall and Spring had the highest R squared values for both the daily and hourly models. One explanation for this is that cyclists are more influenced by the weather in between summer and winter when it is becoming either colder or warmer. This may be due to the fact that Fall and Spring are usually the seasons when people have not yet completely got accustomed to the changing temperatures. Another interesting finding here is that solar radiation is only significant in the Fall. Again, this may be because people have a tendency to prefer warm temperatures when transitioning from summer to winter. Additionally, the beta coefficients values for air temperature in the summer models are negative, whereas the beta coefficient values for air temperature are positive for all other seasons. This suggests that cyclists prefer lower temperatures in the summer, because cycling is uncomfortable in high temperatures. This result is useful for urban planners and designers to design effectively throughout all seasons. For example, designing streets to provide shade and wind flow for cyclists in the summer is important, while in the Fall, streets should allow for sunlight.

Table 8: Model 5 daily regression beta coefficient results

	Winter	***Spring***	***Summer***	***Fall***
Intercept	0.04757	-0.01331	0.01204	0.01882
Residual Ta	0.20405*	0.26994*	-0.29010*	0.64070*
Residual WS	-0.12687	-0.17461	-0.06817	-0.28237*
Residual Solar	0.08744	0.05979	0.13890*	-0.27530*
Residual RH	-0.79567*	-0.51949*	-0.41032*	-0.70713*
Adjusted R squared	0.42380	0.43520	0.34710	0.59160
Model F	7.069	18.53	13.1	30.34

* $p < .05$

Table 9: Model 5 hourly regression beta coefficient results

	Winter	***Spring***	***Summer***	***Fall***
Intercept	0.04177	-0.00513	0.00877	0.01220
Residual Ta	0.07246*	0.24265*	-0.02912	0.37121*
Residual WS	-0.07257	-0.14535*	-0.06916*	-0.15445*
Residual Solar	0.02600	0.00034	-0.00136	-0.02394
Residual RH	-0.75759*	-0.47202*	-0.38483*	-0.61573*
Adjusted R squared	0.08323	0.09549	0.03792	0.09038
Model F	17.32	57.66	22.59	48.49

* $p < .05$

The regression model using UTCI values was statistically significant, which means that outdoor thermal comfort can also be used a predictor for cycling activity. Consistent with the other models, the R squared is higher for daily values. The results of this regression can be used to create an equation that can predict cycling activity based on UTCI values. This is useful for urban planners and designers who wish to increase outdoor activities by changing the built environment. By changing streetscapes and buildings surrounding a street, the thermal comfort will significantly change which would then affect people's decisions to engage in outdoor activities.

Table 10: Model 6 regression beta coefficient results

	Daily	***Hourly***
Intercept	-0.0364291*	-0.0446818*
UTCI	0.0040818*	0.0045439*
Adjusted R squared	0.05169	0.01207
Model F	17.74	90.96

* $p < .05$

CHAPTER 7

SPATIAL ANALYSIS

7.1 Data Processing

The UTCI simulations were conducted for each community district in Manhattan. The first step was to create a framework by using just one of the community districts and then applying that framework (Appendix A) to the rest of the community districts to comprehensively calculate all the hourly UTCI values in 2018 for the Citi Bike stations within Manhattan. The solar radiation map (Appendix B) illustrates how different areas will receive different amounts of solar radiation depending on the direction of the sun as well as the height and orientation of the buildings surrounding the station. When comparing the solar radiation map and the UTCI map, it is evident that they have similar spatial patterns. This similarity demonstrates the strong effect solar radiation has on UTCI values. Furthermore, the effect of wind speed can be seen by comparing the UTCI maps before (Appendix C) and after (Appendix D) implementing the simulated wind speed values (Appendix E) at each station. Stations that experience high wind speeds significantly decreases the UTCI values after implementing the simulated wind speeds. Similarly, the UTCI values do not change significantly in areas that experience low wind speeds. This further indicates the importance of microclimate simulations when examining the relationship between weather and cycling. Microclimate simulations can produce a more accurate representation of the environment, rather than using a single weather station and applying those weather conditions to all bike stations.

After obtaining the UTCI values for every hour of 2018 at each station, the data collected in grasshopper were put back into R and were re-formatted to combine with cycling counts for each station, as well as with spatial data as shown in Table 11. Because numerous stations had NULL values when looking at hourly data, this data was aggregated to daily, weekly, and monthly values as well. A spatial regression model was created to test whether or not there is a spatial aspect to the relationship between UTCI and cycling counts.

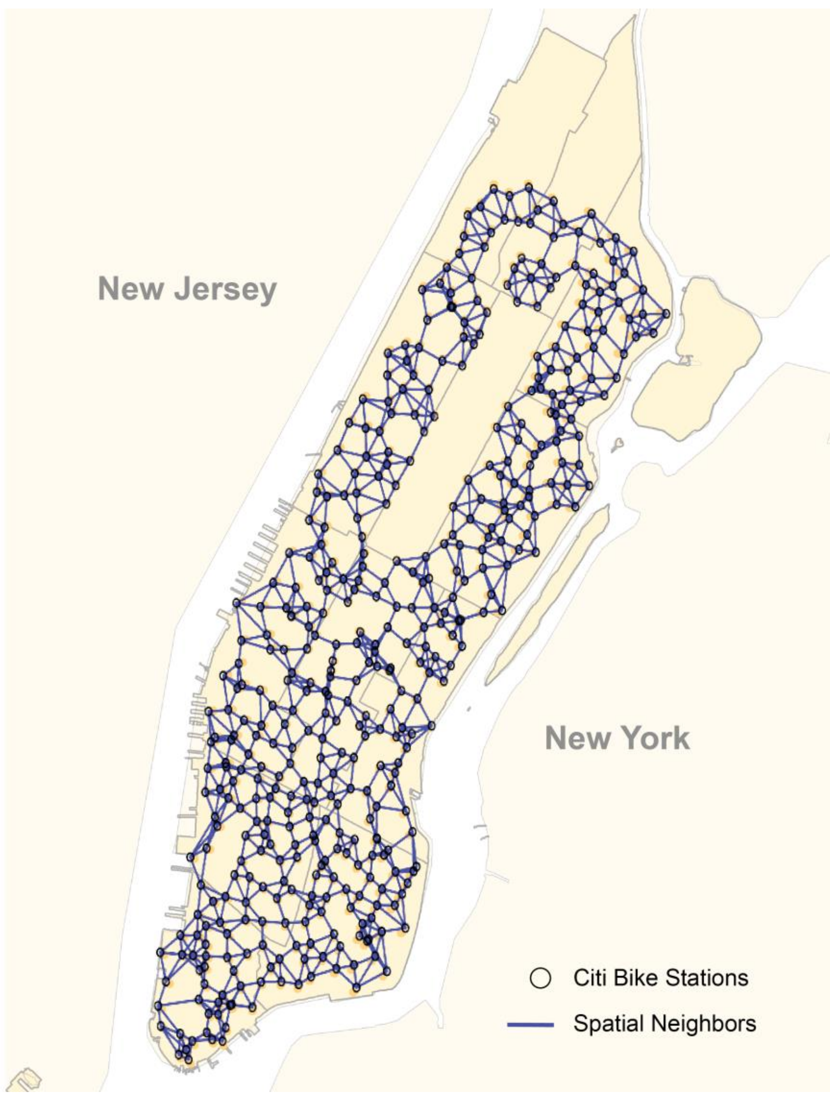
Table 11: Example of monthly data structure from Community District 1

ID	Station Name	Latitude	Longitude	Geometry	UTCI	Count
195	Liberty St & Broadway	40.70906	-74.0104	c(583589.540899192, 4506931.18672383)	14	67
248	Laight St & Hudson St	40.72185	-74.0077	c(583802.919623657, 4508354.38741048)	37	34
249	Harrison St & Hudson St	40.71871	-74.009	c(583698.491568793 , 4508004.16786214)	29	40
257	Lispenard St & Broadway	40.71939	-74.0025	c(584249.050093871, 4508086.1605005)	23	60
259	South St & Whitehall St	40.70122	-74.0123	c(583438.116039422, 4506059.64365467)	36	53
260	Broad St & Bridge St	40.70365	-74.0117	c(583491.194554085, 4506330.07982864)	15	21

7.2 Spatial Regression Modeling

Prior to running the spatial regression models, a linear regression model was used as a base line for the conditional autoregressive (CAR) and simultaneously autoregressive (SAR) models. CAR models show more of a neighborhood effect than SAR models, since in the CAR model, the probability of a certain value is based on neighbor values. The CAR model uses a symmetric weight matrix, which means that directional processes cannot be modeled. In contrast, SAR can model anisotropic models, implying that the processes do not have to be symmetric. These models were used to find the regression coefficient and their standard deviations, as well as the loglikelihood. Spatial regression models require neighborhood lists, in order to test for spatial correlation. Figure 5 shows the resulting neighborhood structure based on the Citi Bike station locations in Manhattan. All modelling was completed using R.

Figure 5: Neighbors Connections



To examine whether a spatial pattern exists in the dataset, the CAR and SAR models were first compared to a non-spatial regression model. Before running the regression models, the data were manipulated to allow easier analysis. Some stations were not being used often, resulting in very few bike trips at the hourly, daily, and weekly levels subsequently making it difficult to see temporal patterns and trends. Therefore, UTCI values per station were aggregated to the monthly level to mitigate this issue. The resulting data contained months as columns and stations as rows, where each value represented the average UTCI value at a particular month and station. After the data manipulation was complete, the linear regression models were created for non-spatial, CAR, and SAR assumptions. The results of the regression models are shown below in Table 12 and 13.

Comparing the loglikelihoods for the three models, the CAR model had the highest average loglikelihood of -1772.33, the next highest was the SAR model with an average loglikelihood of -1772.64, and the non-spatial regression model had the lowest average loglikelihood of -1779.84. The loglikelihood is a measure of the fit of the coefficients. Since the sample sizes were the same for all models, the loglikelihood can be used to determine which model has the best fit for the dataset. The CAR and SAR models resulted in a higher loglikelihood than the non-spatial model, illustrating that there is a spatial aspect to the relationship between UTCI values and bike usage. Although the difference is minimal, the loglikelihood values for the CAR models are higher than the SAR models. Since the CAR model assumes that an area is affected by its neighbors, and not neighbor of neighbors, the model results suggest that bike stations nearby are likely to experience similar microclimate conditions than bike stations at a distance. These results highlight the importance of considering the adjacent areas of the point of interest when simulating location specific UTCI.

The non-spatial regression models are statistically significant for the months of April to October. Similarly, for the CAR and SAR models statistical significance can be seen for the months of April to June, and August to October. This is consistent with the findings from the seasonal model explained earlier, where Fall and Spring had the highest R squared values. Both models confirm that bikers are more likely to be influenced during ‘changing seasons’, where temperatures are unstable, and people are not yet accustomed to the changing temperatures. Another interesting finding is that the magnitude of beta coefficients is much larger for the simulated UTCI models (Models 7, 8, and 9) compared to the non-location specific UTCI model (Model 6). The beta coefficient reflects how severely cycling activity will be affected by UTCI. This suggests that simulated location specific UTCI values have a stronger effect on cycling than the non-location specific UTCI values calculated using weather data for Manhattan. The importance of simulating UTCI using microclimate data is highlighted in this finding.

Table 12: Model 8 non-spatial regression results for all months in 2018

	Beta Coefficient	*Adjusted R squared*	*Loglikelihood*
January	0.5518	-0.0004388	-1553.451
February	0.05376	-0.002482	-1592.712
March	0.8758	0.003375	-1655.717
April	3.118*	0.01871	-1759.416
May	1.962*	0.01185	-1882.401
June	2.6266*	0.01667	-1886.158
July	1.8639*	0.008246	-1892.45
August	2.782*	0.02373	-1933.046
September	1.8652*	0.01131	-1898.189
October	2.5011*	0.01262	-1932.254
November	0.4662	-0.001003	-1735.632
December	0.01569	-0.002461	-1636.619

* $p < .05$

Table 13: Model 9 CAR and SAR regression results for all months in 2018

	CAR Model			*SAR Model*		
	Beta	**AIC**	**Loglikelihood**	**Beta**	**AIC**	**Loglikelihood**
January	0.40021	3096.8	-1544.415	0.4266	3098.1	-1545.04
February	0.021769	3182.4	-1587.193	0.022584	3183	-1587.513
March	0.74552	3304.6	-1648.31	0.76851	3305.4	-1648.71
April	2.907*	3511.4	-1751.68	2.9374*	3511.8	-1751.881
May	1.60206*	3758.7	-1875.354	1.62381*	3759.6	-1875.779
June	2.27138*	3763.8	-1877.914	2.2898*	3764.4	-1878.187
July	1.5537	3772.4	-1882.206	1.56945	3772.7	-1882.326
August	2.50468*	3860.7	-1926.369	2.54324*	3861.2	-1926.614
September	1.71027*	3795.1	-1893.526	1.72959*	3795.3	-1893.648
October	2.0509*	3859.2	-1925.588	2.0875*	3859.5	-1925.775
November	0.32833	3463.9	-1727.973	0.34591	3464.6	-1728.313
December	-0.04858	3263	-1627.488	-0.02342	3263.8	-1627.917

* $p < .05$

CHAPTER 8

DISCUSSION AND CONCLUSION

With the ever-growing availability of open data sources, it is important to understand how big data can be used for both the temporal and spatial exploration of valuable datasets. This study examines the relationship between weather variables and cycling, as well as UTCI and cycling. Previous studies have discussed the effects of weather variability on urban transportation modes, but few studies have conducted an in-depth examination using a variety of temporal scales with the addition of microclimate simulations to calculate location specific UTCI values. This study aims to fill these research gaps.

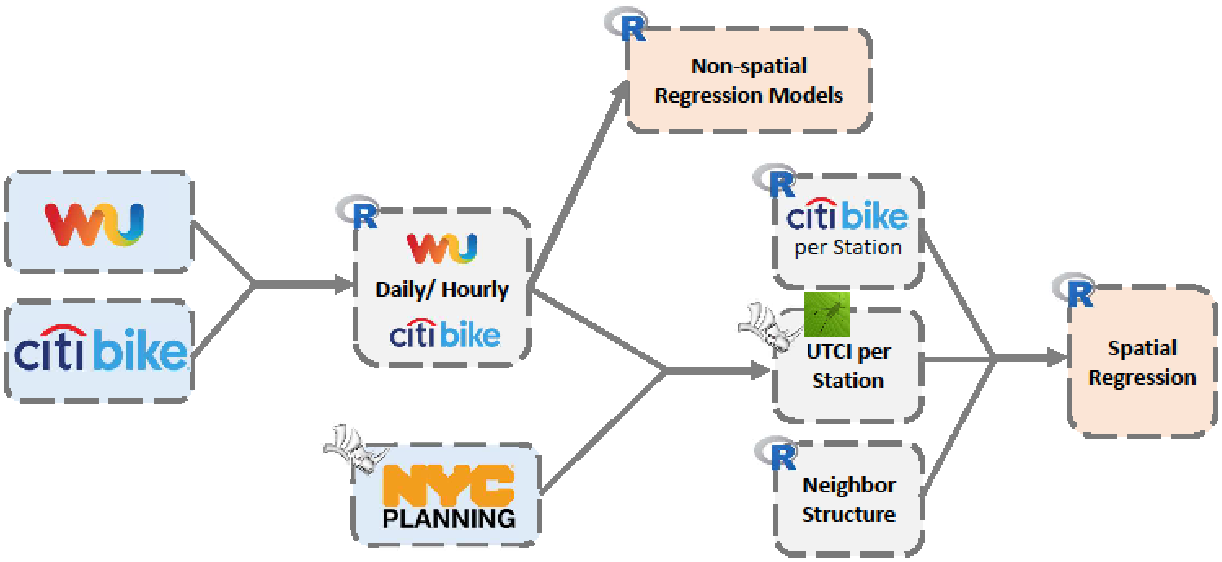
The descriptive statistics indicate that weekday Citi Bike users are mostly commuters, and weekend Citi Bike users are cycling mostly for leisure purposes. In addition, cycling counts are higher during the summer compared to all other seasons, which is partly due to the increase is tourists but also because of the favorable weather. Therefore, different temporal scales were used in the regression models including daily vs hourly, weekday vs weekend, and seasonal effects. The regression models have shown that daily values consistently result in higher R squared values. Weekend cyclists are more likely to be affected by weather variability than weekday cyclists, which is consistent with findings from previous studies (Guo et al., 2007; Kalstein et al., 2009; Singhal et al., 2014). This suggests that weekend cyclists have more flexibility to change transportation modes or shift their cycling time, compared to weekday cyclists who are commuting. Additionally, the regression model examining the seasonal differences demonstrates that cyclists are more likely to be affected by weather variability during the Fall and Spring seasons, which is consistent with a previous study examining seasonal effects (Zhao et al., 2018). The significance of the individual weather variables also varied depending on the season. A negative beta coefficient could be seen for air temperature in the summer model but was positive for all other seasons. In the summer, cyclists prefer a lower temperature for a comfortable cycling experience. The results from the seasonal variation model can be used by planners, policymakers, and urban designers who are seeking to create an optimal outdoor environment throughout all seasons. Finally, the regression model examining the relationship between UTCI and cycling activities was statistically significant, suggesting that UTCI could be used as a predictor for cycling activity.

Results from the microclimate simulations indicate the importance of calculating UTCI values at specific locations when exploring the effects of UTCI on cycling activity. The microclimate simulation allowed for a more accurate representation of the environment in real life, as opposed to using weather values from a single weather station. Wind speeds and solar radiation varied throughout Manhattan due to the built environment, which influenced UTCI values. The UTCI simulation results were used to create spatial regression models to explore the spatial aspects of the relationship between UTCI and cycling activity. A comparison of the non-spatial model, CAR model, and SAR model illustrates that UTCI values are in fact influenced by neighboring UTCI values. Both the CAR and SAR model resulted in higher loglikelihood values, exemplifying the need to consider areas adjacent to the area of interest when examining UTCI. However, the spatial regression models were unable to verify that simulated UTCI values can better predict bike trip counts. This is reasonable since most people do not suddenly decide to ride a bike just because that area is thermally comfortable. Bike users are more likely to decide to ride a bike based on the weather forecast, not the microclimate. Nevertheless, the results of the spatial regression models highlight the importance of further exploration of simulating UTCI values. The simulated UTCI values displayed a stronger effect on bike activity compared to the non-simulated UTCI values, and clearly shows the importance of conducting microclimate simulations. The results of this study can be used to explain why microclimate simulations are crucial in outdoor activity studies.

This study only considered cycling data, but in order to further deepen our understanding of transportation mode choices, it would be useful to look at other transportation mode data and compare results. This would enable a better understanding of the effects of weather variability on transportation mode choice. Due to data availability, this study examined the relationship between weather variables, UTCI and cycling, but this framework can be used with other types of data as well. When temporally and spatially dense data become available, this framework can be used to examine the effects of outdoor environments on pedestrian and outdoor space usage. The exploration of different types of outdoor environments would be beneficial for planners and designers who aim to create an attractive outdoor space. Furthermore, the hourly simulated UTCI values for each station shed light on outdoor thermal comfort throughout the entire Manhattan area. These values can be used to explore different urban design layouts and determine the necessary conditions for promoting an active outdoor area. The regression model results can also assist in creating formulas to predict how much cycling activity would increase, if outdoor thermal comfort is improved. These are all potential avenues for future research.

APPENDIX

APPENDIX A



A diagram showing the workflow of this study

APPENDIX B

Community District 1 Solar Radiation Simulation Results

APPENDIX C

Community District 1 UTCI with Constant Wind Values Simulation Results

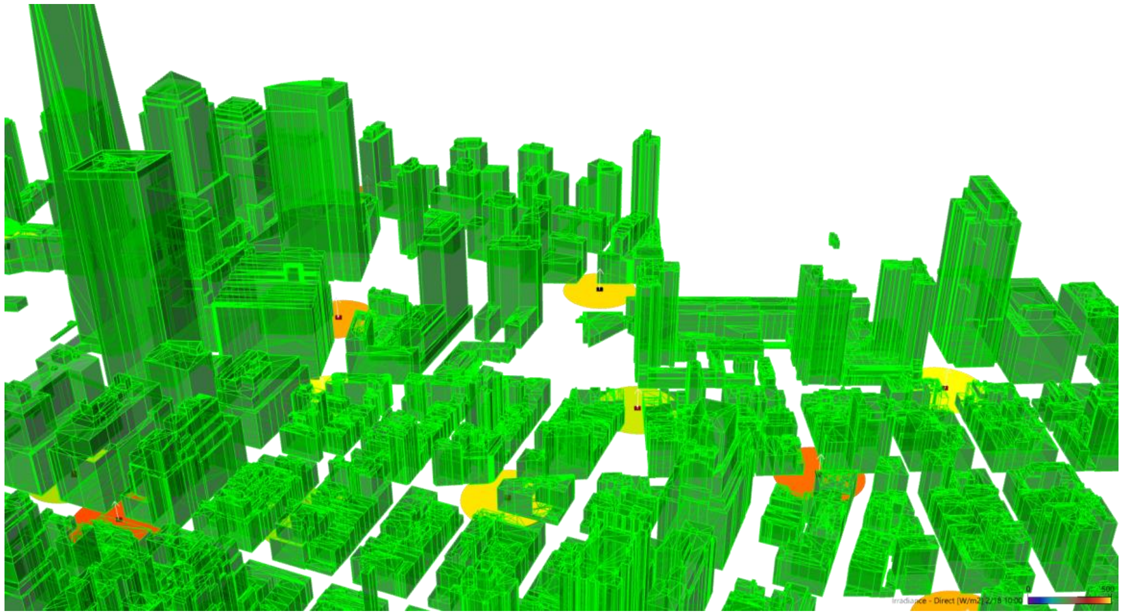
APPENDIX D

Community District 1 UTCI with Simulated Wind Values Simulation Results

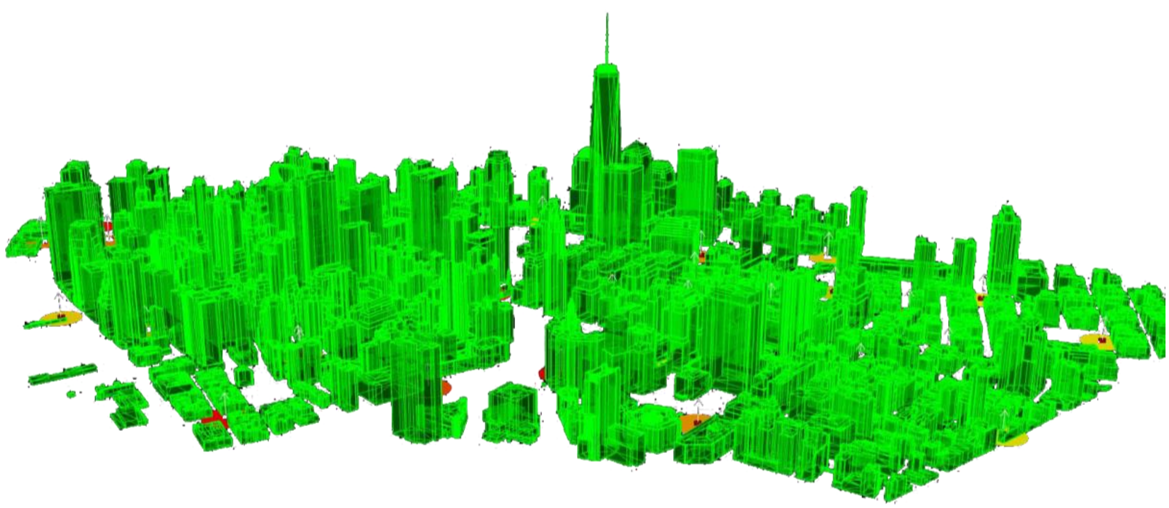
APPENDIX E

Community District 1 Simulated Wind Values Results

APPENDIX F



Perspective view of simulation 3D model (Community District 1)



Zoomed out perspective view of simulation 3D model (Community District 1)

REFERENCES

Bergström, A., & Magnusson, R. (2003). Potential of transferring car trips to bicycle during winter. Transportation Research Part A: Policy and Practice, 37(8), 649-666. doi:10.1016/s0965-8564(03)00012-0

Błazejczyk, K., Jendritzky, G., Bröde, P., & Fiala, D. (2013). An introduction to the Universal thermal climate index (UTCI). Geographia Polonica, 86(1), 5–10. https://doi.org/10.7163/GPol.2013.1

Burchfield, R. A., Fitzhugh, E. C., & Bassett, D. R. (2012). The Association of Trail Use with Weather-Related Factors on an Urban Greenway. Journal of Physical Activity and Health, 9(2), 188-197. doi:10.1123/jpah.9.2.188

Creemers, L., Wets, G., & Cools, M. (2014). Meteorological variation in daily travel behaviour: Evidence from revealed preference data from the Netherlands. Theoretical and Applied Climatology, 120(1-2), 183-194. doi:10.1007/s00704-014-1169-0

Fang, Z., Lin, Z., Mak, C. M., Niu, J., & Tse, K. (2018). Investigation into sensitivities of factors in outdoor thermal comfort indices. Building and Environment, 128, 129-142. doi:10.1016/j.buildenv

Kalkstein, A. J., Kuby, M., Gerrity, D., & Clancy, J. J. (2009). An analysis of air mass effects on rail ridership in three US cities. Journal of Transport Geography, 17(3), 198-207. doi:10.1016/j.jtrangeo

Lai, D., Guo, D., Hou, Y., Lin, C., & Chen, Q. (2014). Studies of outdoor thermal comfort in northern China. Building and Environment, 77, 110-118. doi:10.1016/j.buildenv

Liu, C., Susilo, Y. O., & Karlström, A. (2015). Investigating the impacts of weather variability on individual’s daily activity–travel patterns: A comparison between commuters and non-commuters in Sweden. Transportation Research Part A: Policy and Practice, 82, 47-64. doi:10.1016/j.tra

Miranda-Moreno, L. F., & Nosal, T. (2011). Weather or Not to Cycle. Transportation Research Record: Journal of the Transportation Research Board, 2247(1), 42-52. doi:10.3141/2247-06

Pantavou, K., Theoharatos, G., Santamouris, M., & Asimakopoulos, D. (2013). Outdoor thermal sensation of pedestrians in a Mediterranean climate and a comparison with UTCI. Building and Environment, 66, 82-95. doi:10.1016/j.buildenv

Singhal, A., Kamga, C., & Yazici, A. (2014). Impact of weather on urban transit ridership. Transportation Research Part A: Policy and Practice, 69, 379-391. doi:10.1016/j.tra

Sun, W., & Zhang, L. (2015). City Transit Ridership Forecasting Based on Weather Conditions—Case Study of the Chicago Transit Authority (CTA). Icte 2015. doi:10.1061/9780784479384.224

Thomas, T., Jaarsma, R., & Tutert, B. (2012). Exploring temporal fluctuations of daily cycling demand on Dutch cycle paths: The influence of weather on cycling. Transportation, 40(1), 1-22. doi:10.1007/s11116-012-9398-5

Winters, M., Friesen, M. C., Koehoorn, M., & Teschke, K. (2006). The Impact of Climate on Utilitarian Bicycling: Results of a Canadian Study. Epidemiology, 17(Suppl). doi:10.1097/00001648-200611001-01438

Zhao, J., Wang, J., & Deng, W. (2015). Exploring bikesharing travel time and trip chain by gender and day of the week. Transportation Research Part C: Emerging Technologies, 58, 251-264. doi:10.1016/j.trc

Zhao, J., Wang, J., Xing, Z., Luan, X., & Jiang, Y. (2018). Weather and cycling: Mining big data to have an in-depth understanding of the association of weather variability with cycling on an off-road trail and an on-road bike lane. Transportation Research Part A: Policy and Practice, 111, 119-135. d