



## Modeling outdoor thermal comfort along cycling routes at varying levels of physical accuracy to predict bike ridership in Cambridge, MA

Elizabeth Young <sup>a,\*</sup>, Patrick Kastner <sup>b</sup>, Timur Dogan <sup>b</sup>, Ata Chokhachian <sup>c</sup>, Sarah Mokhtar <sup>a</sup>, Christoph Reinhart <sup>a</sup>

<sup>a</sup> Sustainable Design Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>b</sup> Environmental Systems Lab, Cornell University, Ithaca, NY, USA

<sup>c</sup> Chair of Building Technology and Climate Responsive Design, Technical University of Munich, Munich, Germany



### ARTICLE INFO

#### Keywords:

Biometeorological indices  
Outdoor thermal comfort  
Universal thermal climate index (UTCI)  
Comfort simulations  
Bicycle ridership

### ABSTRACT

The Universal Thermal Climate Index (UTCI) has been linked to outdoor activity patterns and used to evaluate the effectiveness of urban interventions to improve thermal comfort. This study investigates how simulating the urban environment at increasing levels of physical accuracy impacts UTCI values along three cycling routes in Cambridge, Massachusetts. Baseline UTCI values are estimated using a local weather file, and the following increments in physical accuracy are considered: wind-scaling, shading from buildings, shading and cooling from trees, computational fluid dynamics simulations for wind speeds, and simulated surface temperatures. With bike ridership data from Bluebikes, Boston's bike-sharing program, the relationship between bike ridership patterns and UTCI values along each route is studied. Supervised machine learning models are applied to predict bike ridership based on UTCI and other predictors.

UTCI simulation results show that incorporating the various increments of accuracy influences hourly UTCI values at urban areas and exposed areas differently. Incorporating local wind speeds is especially impactful for urban areas. The statistical models trained to predict hourly bike trip counts based on UTCI and other demand and weather predictors achieved a root-mean-squared error of 1.06 trips. 47% of predictions were correct, and an additional 42% of predictions were off by 1 trip.

This study demonstrates the importance of spatial refinement in simulating UTCI, and motivates future research into efficient simulation methods or rules-of-thumb for deriving spatial-temporal UTCI values. Future work into building a robust predictive model would motivate the design of thermally comfortable environments for human-powered transportation in cities.

### 1. Introduction

According to the United Nations, there is a global rush for economic opportunities that attracts people to cities. By 2050, it is expected that 68% of the global population will be urbanites [1]. As of 2020, the transportation sector accounted for 26% of United States greenhouse gas emissions, of which 58% originated from light-duty vehicles [2]. To decarbonize local traffic, a shift towards human-powered transportation such as walking and biking would also improve city dwellers'

psychological and physical health [3], reduce congestion on roads [4], and provide an affordable and comfortable form of transportation for people across all income brackets [5]. Popular planning guidelines for walkable and bikeable neighborhoods are the Complete Streets Projects in the United States [6] and the CIVITAS Initiative by the European Union, which aims to support the European Commission in achieving their sustainable mobility and transportation goals [7]. Additionally, as of April 2021, over 400 cities around the world have introduced initiatives to support walking and biking to adapt to changing demands on

**Abbreviations:** ASHRAE, American Society of Heating, Refrigerating, and Air-Conditioning Engineers; CFD, Computational Fluid Dynamics; MIT, Massachusetts Institute of Technology; MRT, Mean Radiant Temperature; MSE, Mean Squared Error; RMSE, Root-Mean-Squared Error; TIF, Tree Intercepted Fraction; UTCI, Universal Thermal Climate Index.

\* Corresponding author.

E-mail addresses: [elyoung@alum.mit.edu](mailto:elyoung@alum.mit.edu) (E. Young), [pk373@cornell.edu](mailto:pk373@cornell.edu) (P. Kastner), [tkdоган@cornell.edu](mailto:tkdогan@cornell.edu) (T. Dogan), [ata.chokhachian@tum.de](mailto:ata.chokhachian@tum.de) (A. Chokhachian), [smokhtar@mit.edu](mailto:smokhtar@mit.edu) (S. Mokhtar), [creinhart@mit.edu](mailto:creinhart@mit.edu) (C. Reinhart).

<https://doi.org/10.1016/j.buildenv.2021.108577>

Received 14 September 2021; Received in revised form 4 November 2021; Accepted 9 November 2021

Available online 17 November 2021

0360-1323/© 2021 Elsevier Ltd. All rights reserved.

public spaces due to the coronavirus disease 2019 (COVID-19) pandemic [8].

Measures that cities take to this end include introducing a network of bike lanes and widening sidewalks as well as bike sharing services. An additional dimension to human powered transportation is pedestrians' and cyclists' safety and thermal comfort in transit [9]. The latter concern is gaining importance due to rising numbers of heat stress days in cities due to climate change [10]. While there have been efforts to improve outdoor thermal comfort, such as using evaporative coolers and mist-fans for cooling, and tree-planting for shading and cooling, these initiatives have yet to be widely adopted and systematically applied. An essential part of encouraging urban planners and policy makers to promote and implement thermal comfort concepts is the ability to both predict how different urban interventions may impact resident comfort throughout a city over time as well as how those changing thermal sensations may in turn influence mobility mode choices.

Quantifying and evaluating outdoor thermal comfort sensation have been well explored and over 100 biometeorological indices have been developed to-date to measure thermal comfort [11]. Universal Thermal Climate Index (UTCI), which is based on a 187-node model of the human body, incorporates all parameters of the human heat budget and adequately represents the thermal environment. UTCI was shown to be sensitive to changes in ambient conditions such as temperature, solar radiation, wind and humidity [12]. This makes it useful in studying the impact of urban interventions, such as wind breakers or solar shading devices, on local comfort conditions. UTCI is reported in units of degrees Celsius and the comfort level a person experiences at different UTCI values can be assessed through the UTCI assessment scale in Fig. 1.

The relationship between UTCI and people movement has been demonstrated in various applications in the built environment, such as explaining dwelling patterns of regular lunch-goers in an outdoor courtyard [14], evaluating the effectiveness of evaporative coolers in hot arid climates [15], and assessing change in thermal stresses using

shading [16]. A limitation of these aforementioned studies is that they focus on UTCI states at a single location over time. They have, to date, not been used to consider UTCI variations in microclimatic conditions across space such as along a bike path. One may however suspect that an individual's decision to opt for a bike sharing service rather than public transportation or private vehicle is partially influenced by the expected thermal conditions along the route. Extreme cold stress or heat stress and heavy rain are likely to deter even the most committed cyclist.

For a non-spatial annual UTCI analysis, a planner can simply consult a local weather file to calculate hourly UTCI values for an unshaded and exposed position as shown in Fig. 2.

In contrast, a UTCI analysis that considers high resolution microclimatic effects along an urban route is complicated to set up and computationally intensive. There are also different levels of accuracy that can be employed for such a task. The most basic level of spatial accuracy involves predicting whether a location along a route of interest is shaded from sunlight. Additional insight can be gained by considering the effects of seasonal foliage, local wind patterns and exact temperatures of surrounding surfaces such as buildings, ground and vegetation.

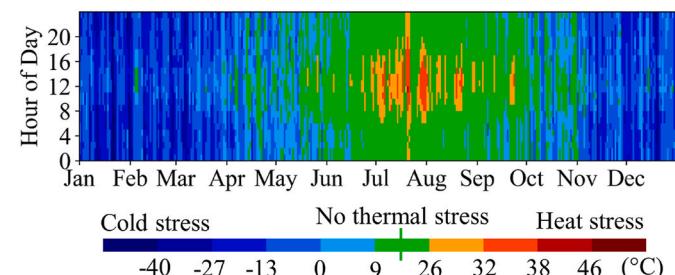
This study seeks to understand how simulating the urban environment at such increasing levels of spatial and physical refinement impacts UTCI values along three bike routes in Cambridge, Massachusetts. Six variants of UTCI calculations are generated for multiple years along these routes. The first variant uses only inputs from a local weather file. Subsequent variants incrementally consider the following elements: 1) Uniform scaling of wind speeds for urban wind shading, 2) shading from buildings, 3) shading from trees, 4) spatially resolved wind conditions due to building geometry, and 5) surface temperatures of the environment along the route. Tools used for the UTCI calculations include DIVA-for-Rhino for solar radiation simulations [17], PANDO for modeling tree seasonality and leaf temperatures [18], Eddy3D for Grasshopper for incorporating Computational Fluid Dynamics (CFD) wind results [19], and Surfer for simulating building and ground surface temperatures [20].

The outcomes of this analysis can guide urban planners and their environmental consultants on the level of simulation accuracy required to adequately capture UTCI comfort levels in different urban settings. In addition, the study investigates the relationship between UTCI along the three bike routes in Cambridge and bike ridership along these routes, which is determined from publicly accessible bike trip data from Boston's Bluebikes program [21]. With the Bluebikes data and six sets of simulated UTCI values, statistical models are trained to predict bike ridership based on a set of environmental conditions. If successful, such a predictive model would be a step towards enabling decision makers to quantify the effect of modified thermal comfort on local bike ridership.

## 2. Methodology

### 2.1. Site description

UTCI is simulated along three bike routes from the Bluebikes bike sharing program in the metropolitan area in Boston, with two out of



**Fig. 2.** Annual heatmap of hourly UTCI in Boston (2019) as calculated from weather file data.

**Fig. 1.** UTCI Assessment Scale, adapted from UTCI official page [13].

three routes located at the Massachusetts Institute of Technology (MIT) campus. Boston has a hot summer continental climate (Köppen -Geiger classification is Dfa), with hot summers and very cold winters. Three station pairs are chosen for this study (Fig. 3) based on the following characteristics: 1) the travel route between each station pair is identifiable as there are few to no alternative routes that cyclists would take, 2) the routes present unique environmental conditions and variations, and 3) the routes are among the station pairs with the highest number of bike trips.

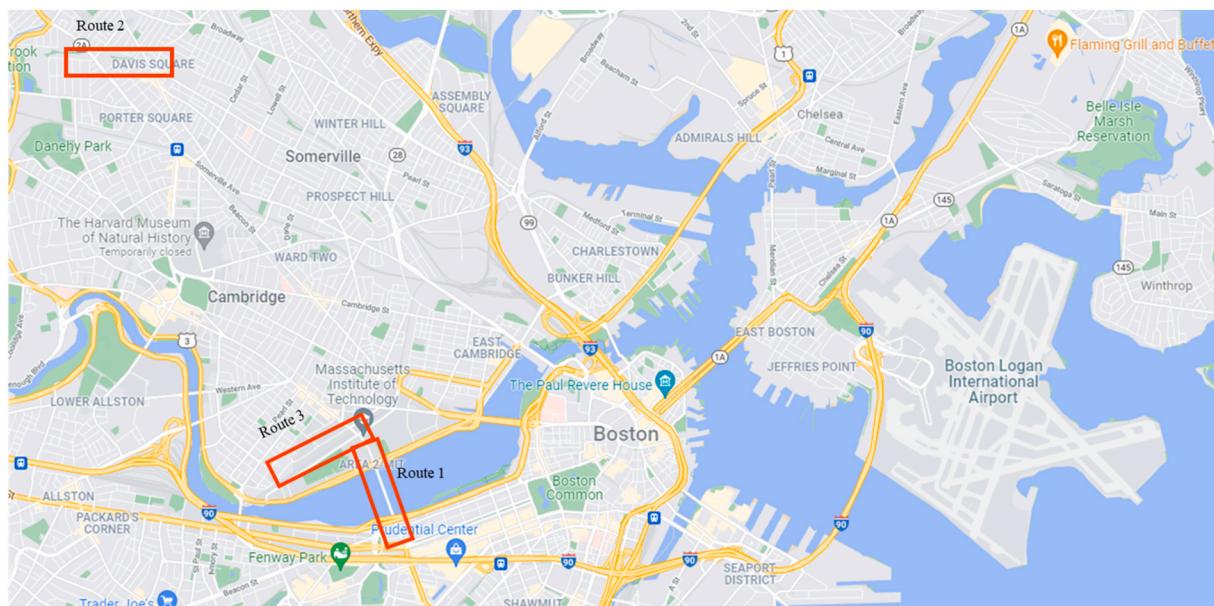
(Imagery ©2021 Google, Imagery ©2021 CNES/Airbus, MassGIS, Commonwealth of Massachusetts EOEA, Maxar Technologies, Sanborn, USDA Farm Service Agency, Map data ©2021 Google).

The first route is across a highly exposed bridge, between Bluebikes station “MIT at Mass Ave/Amherst Street”, which is located in MIT’s main campus, and “Beacon St/Mass Ave” across the Charles River in Boston. The second route crosses a linear park with trees, between “Linear Park – Mass. Ave at Cameron Ave” and “Davis Square”. The third route is between “MIT Pacific St at Purrington St”, which is located in front of MIT’s largest graduate apartment complex, and “MIT at Mass Ave/Amherst Street” on MIT’s main campus. While there are a few ways

to ride between these stations, the safest and most probable route is along Vassar Street as shown Fig. 3.

## 2.2. UTCI simulations

Hourly UTCI is calculated along each route for three years, between 2017 and 2019. The equation used to calculate UTCI is a Python version of the original UTCI approximation written in Fortran [13] and requires four meteorological inputs: 1) air temperature [ $^{\circ}\text{C}$ ], 2) relative humidity [%], 3) mean radiant temperature [ $^{\circ}\text{C}$ ], and 4) wind speed [m/s] at 10 m above ground. Six variations of UTCI calculation methods are defined. With each subsequent variation, the level of spatial resolution and complexity in calculating mean radiant temperature (MRT) and wind speed are increased. For all variants, the hourly air temperature and relative humidity are sourced from historical weather data from Boston-Logan International Airport’s weather station, which are provided courtesy of Dr. Drury B. Crawley and Linda Lawrie [22]. A summary of the data sources for all four meteorological inputs is provided in Table 1.



**Route 1:**

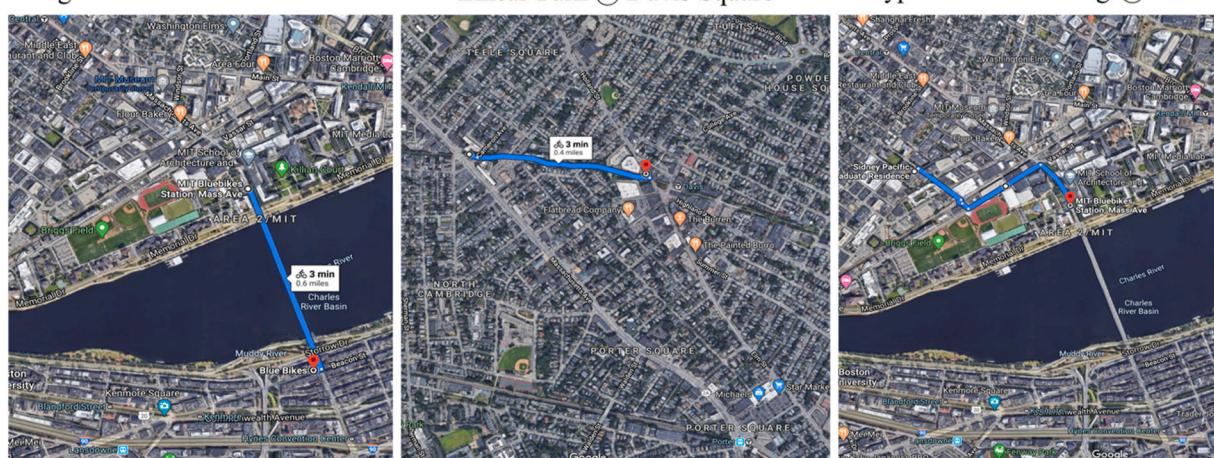
Bridge between MIT and Beacon St

**Route 2:**

Linear Park @ Davis Square

**Route 3:**

Typical Urban Setting @ MIT



**Fig. 3.** Maps of Boston area and three Bluebikes routes.

**Table 1**

Summary of meteorological inputs for variant 1 to variant 6.

Variant	Data Source for Meteorological Inputs				UTCI output
	Wind Speed [m/s]	Mean Radiant Temperature [°C]	Air Temperature [°C]	Relative Humidity [%]	
1: Weather Station	Weather station	Weather station	Weather station	Weather station	1 UTCI value every hour
2: Wind-scaling	Weather station (urban scaled)	Weather station	Weather station	Weather station	1 UTCI value every hour
3: Buildings	Weather station (urban scaled)	DIVA simulation of solar radiation with buildings	Weather station	Weather station	UTCI for each hour along route at 10 m intervals
4: Trees	Weather station (urban scaled)	DIVA simulation of solar radiation with buildings and trees	Weather station	Weather station	UTCI for each hour along route at 10 m intervals
5: CFD	CFD simulation using Eddy3D	DIVA simulation of solar radiation with buildings and trees	Weather station	Weather station	UTCI for each hour along route at 10 m intervals
6: Surface Temperatures	CFD simulation using Eddy3D	Simulated building and ground surface temperatures using Surfer, leaf temperatures and phenology using PANDO	Weather station	Weather station	UTCI for each hour along route at 10 m intervals

### 2.2.1. Variant 1: weather station

Variant 1 calculates UTCI from hourly historical weather data from Boston-Logan International Airport's weather station from years 2017–2019, requiring minimal simulation effort and data preparation. Variant 1 assumes an environment that is highly exposed to the ground and sky. Hourly values of air temperature, relative humidity and wind speed at the weather station are directly provided by the weather file and used to calculate UTCI. The weather station at Boston-Logan International Airport is 6 km–9 km away from the three cycling routes. MRT is derived from other data parameters provided by the weather file.

MRT calculations in Variant 1 through 4 closely follow the methods presented by Kessling et al. [23], which calculates total MRT from long wave MRT, short wave diffuse MRT, and short wave direct MRT. Equation (1), Equation (7), and Equation (8) are directly adapted from the Kessling paper.

In Equation (1), long wave MRT is calculated from the long wave radiation of n isothermal surfaces with emission coefficient  $\epsilon$ , temperature T, and view factor F.

$$MRT_{lw} [\text{°C}] = \left[ \sum_{i=1}^n \epsilon_i T_i^4 F_i \right]^{\frac{1}{4}} \quad (1)$$

In Variant 1, the emission coefficient of all surfaces is 0.95. Two distinct surfaces are considered – the sky and the ground – and each surface takes up 50% of the total view factor. The ground temperature is equal to the air temperature from the weather file, and the sky temperature is derived using Equation (2), Equation (3), and Equation (4). While the MRT calculation for variant 1 through 4 closely follows the Kessling method mentioned above, it deviates in its calculation of sky temperature,  $T_{sky}$ . Instead of adopting the clear sky condition assumption in the Kessling method, this study accounts for cloud cover in Equation (2), Equation (3), and Equation (4) by adopting EnergyPlus' Sky Temperature Calculation method. No extra effort in simulation is required as cloud cover is readily available in the historic weather files for Boston-Logan International Airport's weather station.

The Stefan-Boltzmann constant,  $\sigma = 5.6697 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-4}$ , is used in Equation (2) and Equation (3). Cloud cover, N, is taken from the weather file and used in Equation (4).  $T_{drybulb}$  is the air temperature from the weather file, and  $T_{dewpoint}$  is the dewpoint, calculated from Equation (5) and Equation (6) [24].

$$T_{sky} [\text{°C}] = \left( \frac{Horizontal_{IR}}{\sigma} \right)^{1/4} - 273.15 \quad (2)$$

$$Emissivity_{IR} = Emissivity_{sky} \times \sigma \times T_{drybulb}^4 \quad (3)$$

$$Emissivity_{sky} = \left( 0.787 + 0.764 \ln \frac{T_{dewpoint}}{273.15} \right) \times (1 + 0.0224N - 0.0035N^2 + 0.00028N^3) \quad (4)$$

$$y = \ln \frac{RH}{100} + 17.67 \times \frac{T_{ambient}}{243.5 + T_{ambient}} \quad (5)$$

$$T_{dewpoint} = \frac{243.5y}{17.67 - y} \quad (6)$$

The long wave MRT equation can thus be simplified to Equation (7) for Variant 1.

$$MRT_{lw} = \left( 0.95 \left[ 0.5T_{sky}^4 + 0.5T_{ambient}^4 \right] \right)^{\frac{1}{4}} \quad (7)$$

In Equation (8), short wave diffuse MRT is calculated from the diffused short-wave radiation,  $D_i$ , emitted from n surfaces with view factor F.  $\epsilon_p$ , the emission coefficient of clothing or skin, is 0.95.  $\sigma$  is the Stefan-Boltzmann constant.  $a_s$ , the short-wave absorption coefficient of a person, is the product of the clothes solar absorption that is 0.5 and the fraction of the body exposed to environmental radiation that is 0.725 for a standing person [25].

$$MRT_{diff} = \left[ \sum_{i=1}^n a_s \frac{D_i}{\epsilon_p \cdot \sigma} F_i \right]^{\frac{1}{4}} \quad (8)$$

In Variant 1, the diffuse radiation incident on the person comes from the sky and the ground. The diffuse solar radiation from the sky is obtained directly from the weather file and the diffused radiation from the ground is a product of the global horizontal irradiance incident on the ground and the ground reflectance value of 0.2. The short wave diffuse MRT can thus be simplified to Equation (9) for Variant 1.

$$MRT_{diff} = \left[ [0.5DiffRad_{sky} + 0.5DiffRad_{ground}] \frac{a_s}{\epsilon_p \sigma} \right]^{\frac{1}{4}} \quad (9)$$

In Equation (10), the short wave direct MRT is calculated from the radiation intensity of the sun,  $I$ .  $f_p$  is the projected area factor which is the fraction of the human body that is exposed to direct solar radiation and depends on the elevation of the sun.

$$MRT_{dir} = \left[ f_p a_s \frac{I}{\epsilon_p \sigma} \right]^{\frac{1}{4}} \quad (10)$$

In variant 1, the direct radiation incident on the person is obtained directly from the weather file.

The total MRT is calculated with Equation (11) and used as an input into the hourly UTCI calculation alongside air temperature, relative humidity, and wind speed.

$$MRT = \left[ MRT_{lw}^4 + MRT_{diff}^4 + MRT_{dir}^4 \right]^{\frac{1}{4}} \quad (11)$$

### 2.2.2. Variant 2: wind-scaling

Variant 2 calculates UTCI by additionally scaling wind speeds from the weather station to account for difference in terrain between the weather station and the site of interest. Air temperature, relative humidity and MRT inputs are equal to those in Variant 1, while the weather station wind speed is modified by Equation (12), which was directly adapted from American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) 2001 Fundamentals Handbook.

$$WSp_{urban} = WSp_{epw} \left( \frac{\delta_{epw}}{z_{epw}} \right)^{a_{epw}} \left( \frac{z_{urban}}{\delta_{urban}} \right)^{a_{urban}} \quad (12)$$

With the three routes located in an urban and metropolitan area, and the weather station located in the airport, according to the terrain dependent coefficients in ASHRAE 2001, the following values apply:  $\delta_{epw} = 270\text{m}$ ,  $a_{epw} = 0.14$ ,  $\delta_{urban} = 460\text{m}$  and  $a_{urban} = 0.33$ .  $z_{epw}$  and  $z_{urban}$  are 10 m. Equation (12) is thus simplified to Equation (13).

$$WSp_{urban} = 0.45 WSp_{epw} \quad (13)$$

### 2.2.3. Variant 3: buildings

Variant 3 calculates UTCI that varies along a given route as the MRT input is varied spatially based on buildings along the routes. Geometric models of the three routes and their surroundings are assembled in Rhinoceros 3D from Geographic Information System (GIS) building base maps and their accompanying building height attributes. Base maps from the Cities of Boston, Cambridge and Somerville are used. For the City of Somerville, building heights are unavailable and are estimated to be 10.7 m (or 35 feet high) for typical residential buildings and 14.6 m (48 feet high) for slightly taller commercial buildings. The buildings are modeled on a flat terrain at ground level as the Boston area is relatively flat.

Air temperature, relative humidity, and wind speed inputs are equal to those from Variant 2. MRT is calculated using a similar method to Variant 1 and 2, but accounts for buildings in addition to the default sky and ground. With building geometries varying along the routes, the view

factors of the sky and other geometry varies as well. Building view factors (Fig. 4) are higher in urban areas (probe 0–9 and 81–89) and lower in exposed areas such as the river (probe 10–80). Small oscillations in view factors are present in Fig. 4 due to discrete surface subdivisions in the geometry meshing.

To calculate long wave MRT for Variant 3, the temperature of any non-sky surfaces is assumed to be equal to the air temperature. The sky temperature is equal to that in Variant 1. The view factor of all other surrounding surfaces is  $1 - VF_{sky}$ . The long wave MRT at each probing point along a route is calculated using Equation (14).

$$MRT_{lw(i)} = (0.95 [VF_{sky(i)} T_{sky} + (1 - VF_{sky(i)}) T_{ambient}])^{\frac{1}{4}} \quad (14)$$

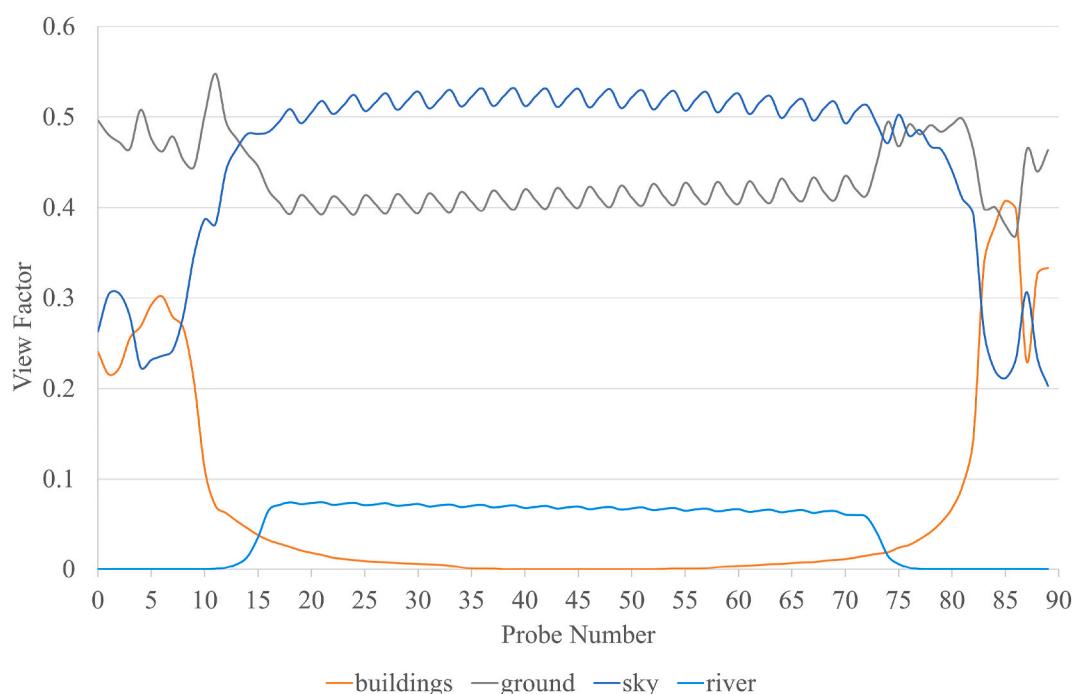
The short wave direct MRT and short wave diffuse MRT are also geometry-dependent. While the direct and diffuse solar radiation in Variant 1 are obtained directly from the weather file and do not vary along the routes, these values for Variant 3 are simulated using DIVA-for-Rhino's radiation map component and thus vary along the routes. The DIVA simulations are run using the DAYSIM implementation, a RADIANCE-based daylighting analysis software that models the annual amount of daylight in and around buildings [26] to simulate hourly direct and diffused solar radiation on a 5m by 5m grid at a height of 1.6 m above ground. All buildings have the material "OutsideFacade\_30", which is one of the default outdoor façade materials from the DIVA materials library.

The simulated solar radiation values are then used to calculate diffuse MRT and direct MRT along the route using Equation (9) and Equation (10) respectively. Total MRT is calculated using Equation (11).

### 2.2.4. Variant 4: trees

Variant 4 builds on Variant 3 by further considering shading from the trees when simulating direct and diffuse solar radiation. Location of trees along each route are obtained through the Cambridge Tree Inventory [27], the City of Somerville's Tree Inventory [28] and Google Maps [29]. The trees are categorized into three groups based on size, with dimensions summarized in Table 2. PANDO, a Grasshopper plugin tool for modeling trees is used to create the tree geometries.

The radiation simulations assume that the trees are completely



**Fig. 4.** View factors of buildings, ground, sky and river along route 1, with probe 0 at MIT/Amherst ST and probe 89 at Beacon St.

**Table 2**

Dimensions of Trees assumed in simulations.

Size	Radius of Crown [m]	Crown Height [m]	Trunk Diameter [m]	Trunk Height [m]
Small	2.7	4	0.5	4
Medium	4.7	8	0.5	8
Large	6.5	11	0.5	14

DIVA simulations with the same assumptions as Variant 3 are run with the buildings and tree geometries. With the simulated diffuse and direct radiation values, MRT is calculated with the same methods as Variant 3. When calculating long-wave MRT, view factors are calculated from a model that includes trees. All non-sky surfaces, including trees, have surface temperatures equal to air temperature.

opaque year-round. To account for tree seasonality and porosity, an interpolation between the MRT values without trees (from Variant 3) and MRT values with trees is performed in Equation (15).

$$MRT_{result} = TIF(MRT_{notrees}) + (1 - TIF)(MRT_{withtrees}) \quad (15)$$

The factor used to interpolate between MRT values in Equation (15) is the Tree Intercepted Fraction (TIF), which is the fraction of solar radiation passing through a tree in a given hour [30]. The TIFs range from 0.7 in summer to 1 in winter. The TIFs, calculated in PANDO, are calculated using Equation (16).  $K_{bs}$ , the extinction coefficient of the leaves, is 0.5.  $\Omega$ , is the clumping factor of leaves depending on the tree species.  $W, X$  is the distance between trees and the crown size respectively. LAD, referring to leaf area density, and the clumping factor are defined as having values of 1.

$$TIF = 1 - \exp\left(-K_{bs}\Omega LAD \frac{W}{X}\right) \quad (16)$$

#### 2.2.5. Variant 5: CFD wind

Variant 5 builds on Variant 4 by considering the buildings and trees in the vicinity of the routes when calculating wind speeds. Air temperature, relative humidity and MRT inputs are identical to Variant 4, while wind speeds vary along the route based on computational fluid dynamics (CFD) simulations. Eddy3D, a Grasshopper plugin for airflow and microclimate simulations [19], is used to set-up and run CFD simulations in BlueCFD/OpenFOAM, a widely-used open source tool for CFD simulations [31]. The extent of the geometric models used in the CFD simulations are shown in Fig. 5.

Table 3 summarizes the assumptions of the CFD simulations.

Two CFD simulations are run for each route – one with trees using the Darcy-Forchheimer method and one without trees. The wind velocity field of the simulations are probed at a height of 1.6 m and at 10 m intervals along each route, which ensures that they are at the same location as the MRT calculations.

To account for tree porosity and seasonality, TIFs are used to interpolate between wind speeds with trees and wind speeds without trees in Equation (17).

**Table 3**

Summary of CFD settings for Variant 5.

Category	Variable	Values for Route 1 & 3	Values for Route 2
Boundary Conditions (Atmospheric Boundary Layer Flow)	Wind Directions	0,45,90,135,180,225, 270,315	0,45,90,135,180,225, 270,315
	Reference Velocity, URef	5 m/s	5 m/s
	Reference Height, ZRef	10 m	10 m
	Surface Roughness length, $Z_0$	2 m, representing an urban area	2 m, representing an urban area
	Minimum Z-coordinate, zGround	0 m	0 m
	Block Size	30 m	30 m
	Inner Rectangle Size	1000 m	600 m
	Outer Radius Size	3600 m	2300 m
	Height	Best practice value (default)	Best practice value (default)
	Accuracy of Building Mesh	3	3
Simulation Domain (Cylindrical)	Accuracy of Features	3	4
	Accuracy of Bounding Box	0	0
	Accuracy of Ground Mode	3	4
	Number of Iterations	With snapping, no layers	With snapping, no layers
	Turbulence Model	kEpsilon	kEpsilon
	Relaxation Factors	Optimized	Optimized
	Solution and Algorithm Control	Optimized	Optimized
Run Settings	Number of Iterations	2000	2000
	Turbulence Model	kEpsilon	kEpsilon
	Relaxation Factors	Optimized	Optimized
	Solution and Algorithm Control	Optimized	Optimized

$$WSp_{Variant4} = TIF(WSp_{notrees}) + (1 - TIF)(WSp_{withtrees}) \quad (17)$$

These wind speeds are probed at a height of 1.6 m while UTCI requires input wind speeds at a height of 10 m. While probing at a height of 10 m could be an option, the aim of the study is to determine the effect of geometry on thermal comfort at a human height. Therefore, wind is probed at a human height of 1.6 m and scaled-up after. Equation (18) is used to scale-up the wind speeds [32].

**Fig. 5.** Extents of CFD simulations for route 1 and 3 (left) and route 2 (right).

$$WSp_{10m} = WSp_{1.6m} \frac{\ln\left(\frac{10+0.01}{0.01}\right)}{\ln\left(\frac{1.6+0.01}{0.01}\right)} \quad (18)$$

### 2.2.6. Variant 6: surface temperatures

In Variant 6, surface temperatures of various urban features are considered when calculating MRT. The air temperature, relative humidity, and wind speed inputs are identical to Variant 5. Hourly leaf surface temperatures are simulated in PANDO and assigned uniformly to all trees in the model. The average difference between leaf surface temperature and air temperature is approximately 0.5 °C for 2017, 2018, and 2019.

Building and ground surface temperatures are simulated with Surfer, an algorithm used to predict exterior surface temperatures in complex urban 3D models [20]. The algorithm discretizes the model surfaces into mesh cells and utilizes K-Means clustering to group these cells based on material properties and exposure to the sun and sky. In this study, the ground is assigned an asphalt material and the building surfaces are assigned a red brick material which is typical in Boston. The surfaces are split into ten clusters based on the surface solar absorptance, thermal absorptance, conductivity, density specific heat, orientation and construction material. Five clusters are for the ground and five are for the buildings. Representative surface patches for each cluster are exported and a shoebox is placed at the location of each representative surface patch to prepare for the EnergyPlus simulation through ClimateStudio [33]. The EnergyPlus simulation produces hourly surface temperatures of the representative patches and hourly sky temperatures. The hourly surface temperatures of the representative patches are then assigned to each mesh cell in the model based on the clusters they belong to. At each probing point along the route (at 10 m intervals and 1.6 m above ground), the MRT due to the buildings and ground is calculated from 1) the surface temperatures of each mesh cell, and 2) the view factors of each mesh cell from the probing point. The aggregate hourly temperatures,  $T_{building+ground}$ , along the route are later used in the total MRT calculation.

In Route 1's case where a river is present over a large portion of the route, hourly river surface temperatures are estimated as the 7-day moving average of the weather file's hourly air temperature as actual river temperatures are only available from May to October each year. Nonetheless, these measurements were used to validate the 7-day moving average estimate. Annual root mean squared error (RMSE) between the moving average estimations and the actual river temperatures range from 2.1 °C to 2.6 °C, while the RMSE between the default air temperature and the actual river temperature range from 3.9 °C to 4.6 °C. Where the Harvard Bridge crosses the Charles River, the river surface and the bridge surface were both modeled at ground level.

The effect of solar radiation on MRT closely follows the methods presented in Dogan et al. (2021). It is represented as a temperature offset, "dMRT", that is calculated using the Effective Radiant Field [34] in an outdoor setting.

With the dMRT that accounts for short-wave radiation, and the various surface temperatures and view factors that account for long-wave radiation, the total MRT is calculated using Equation (19).

$$\begin{aligned} TotalMRT &= dMRT \\ &+ VF_{buildings+ground} T_{building+ground} + VF_{trees} T_{trees} \\ &+ VF_{sky} T_{sky} + VF_{river} T_{river} \end{aligned} \quad (19)$$

Variant 6's MRT workflow is used to calculate MRT with trees and without trees for each route, and the final MRT, which accounts for tree porosity and seasonality, is calculated using Equation (15).

### 2.3. Statistical analysis

To explore the relationship between UTCI and bike ridership, the correlation between the UTCI obtained from the simulations and the

hourly bike trip count along each route is studied. Furthermore, to explore the contribution of UTCI to a person's decision to bike, the simulated UTCI values are also used as an input to a predictive model for bike ridership along each route.

#### 2.3.1. Bluebikes dataset

The dataset for bike ridership from 2017 to 2019 is obtained from Bluebikes, which includes the following details about each trip: trip duration, start time, end time, start station details (id, name, latitude, longitude), end station details (id, name latitude, longitude), bike ID, user type, birth year and gender. Table 4 lists the number of trips recorded for the routes corresponding to the UTCI simulations.

The number of trips taken along each route for every hour was calculated. These hourly trip counts were used to relate hourly UTCI, among other contributing predictors, to bike ridership. Only trips taken by Bluebikes subscribers, as opposed to pay-as-you-go users, are counted to reduce variability in travel patterns.

#### 2.3.2. Ridership predictors

Aside from UTCI, three other categories of factors that could potentially affect a person's decision to cycle are considered: supply predictors, demand predictors, and weather predictors.

Supply predictors refer to factors that could affect a potential rider's ability to rent a bike when they want to or need to (i.e. the availability of bicycles and docking spaces). The 'Dockability' and 'Releasability' data (together referred to as bike availability data) for the start and end stations were obtained directly from Bluebikes. 'Dockability' (over a given time period) is the percentage of minutes in which the station had at least 15% of its docks available for riders to return bicycles while 'Releasability' (over a given time period) is the percentage of minutes in which the station had bicycles available in at least 10% of its docks.

Bike availability data is provided at 30 min time intervals from 2017 to 2019. Apart from indicating bike and dock availability, the dataset also indicates when bike stations were closed or hibernating for the season. The hourly 'Dockability' and hourly 'Releasability' are calculated by summing the half-hourly values and halving the sum. In studying the relationships between bike ridership and various predictors, hours with 0% 'Dockability' and 'Releasability' are excluded from the dataset.

Demand predictors refer to factors that affect whether a person needs a bike at a given hour. This is largely based on the activity patterns of a person. Given that a single route would attract multiple types of riders with different behavioral profiles, an ideal study would account for individual travel patterns through anonymous cyclist identification numbers. This way, rider-types can be formed with categories such as students, working professionals, and visitors, among others. However, as this data was not available due to privacy concerns, the hourly demand for bikes is encapsulated in the following parameters: the route of interest, the hour of the day, whether the hour is on a weekday or

**Table 4**

Summary of routes (in both directions) and number of trips taken along each route from 2017 to 2019.

Route #	Start Station	End Station	Number of Trips
1a	Beacon St at Massachusetts Ave	MIT at Mass Ave/Amherst St	13,678
1b	MIT at Mass Ave/Amherst St	Beacon St at Massachusetts Ave	13,292
2a	Linear Park - Mass. Ave. at Cameron Ave.	Davis Square	8,508
2b	Davis Square	Linear Park - Mass. Ave. at Cameron Ave.	8,162
3a	MIT Pacific St at Purrington St	MIT at Mass Ave/Amherst St	9,078
3b	MIT at Mass Ave/Amherst St	MIT Pacific St at Purrington St	7,903

weekend, and whether the hour is on a ‘special date’. Special date refers to MIT holidays and public holidays for Route 1a,1b,3a, and 3b, and public holidays for Route 2a and 2b. When MIT is not in regular session during the summer and winter breaks, it is considered a special date. Other special dates, such as spring vacation and other school-related holidays were obtained from the MIT academic calendars [35].

Weather predictors refer to factors that affect the environment along the route. Many meteorological factors are available in a weather file to describe the weather at a given hour – air temperature, wind speed, relative humidity, precipitation, cloud cover, direct and diffuse solar radiation, among others. However, many of these are accounted for in the calculation of UTCI. Hence, the two weather predictors used in this study are precipitation and simulated UTCI. Precipitation is obtained from the Boston-Logan International Airport historic weather file that was used in the UTCI simulations.

To use UTCI from Variants 3 to 6 as a predictor, the spatially resolved UTCI along a route is condensed into a single-value metric, “effective route UTCI”, that quantifies the comfort conditions along a route for a given hour. Three ways of deriving effective route UTCI from spatially resolved UTCI data are explored. The first method, ‘average UTCI’, is an average of the UTCI values across all probing points for a given hour, which assumes that the cyclist is familiar with the thermal comfort conditions along the whole route. The second method, ‘start UTCI’, is an average of the UTCIs at the first three probing points of a route. This assumes that a cyclist would perceive the route conditions by evaluating the conditions at its starting point. In this study, this would be the first 30 m of travel from the starting Bluebikes station. Lastly, the third method, ‘worst segment UTCI’, is an average of the UTCIs at the most thermally uncomfortable segment of the route. The segment spans 10% of the route’s total length. This metric assumes that cyclists are sufficiently familiar with their route to predict comfort conditions during the worst part of a route.

### 2.3.3. Statistical models

To study the relationship between bike ridership and the various supply, demand and weather predictors, the various hourly datasets are combined to form a dataset comprising the variables listed in Table 5.

A statistical model is trained to predict the number of expected bike riders along routes for a given hour using various predictors that affect bike ridership. The full dataset of hourly counts of bike trips with the corresponding demand and weather predictors has a total of 117,620 data points. These data points span all routes and include trips from all hours in a year. As the demand for bikes in non-peak and night hours are likely to be zero, and the proportion of night and non-peak hours are significant, 71% of all data points in the full dataset have zero trips. To build a predictive model that would more accurately predict positive numbers of trips and hence more accurately quantifying demand for bikes, the statistical models in this study are trained on data from hours between 2pm and 6pm, when bike trips were more frequently taken. This narrowed dataset comprises a total of 20,802 data points, where the

**Table 5**

Summary of fields in combined dataset for bike ridership analysis.

Variable	Possible Values
Number of Bike Trips	0 or more
Route Number	1a, 1b, 2a, 2b, 3a, 3b
Year	2017, 2018, 2019
Month	1–12
Hour	0–23
Special Date	1 (special date) 0 (normal date)
Weekend	1 (weekend) 0 (weekday)
Precipitation	0 or more inches of rainfall
UTCI (Total 14 sets from 6 variants and 3 effective UTCI metrics)	–40 °C–46 °C

proportion of zero-trip data points decreases significantly to 55.1%.

The statistical models are trained using Scikit-learn, a Python module for machine learning. The dataset is split into a train set with 80% of the data, and a test set with 20% of the data. A baseline model is used to evaluate the statistical models. The baseline model calculates the average trip count of all data points from the train set and uses this value as the model prediction. The baseline model’s predicted value is 0.82 trips to compare between models is the root mean squared error (RMSE) of the test set (Equation (20)). The RMSE measures the error between a model’s predicted trip counts,  $Y_{pred}$ , and the actual trip counts,  $Y_{actual}$ .

$$RMSE = \sqrt{\sum_{n=1}^{N} (y_{pred}^{(i)} - y_{actual}^{(i)})^2} \text{ for } N \text{ data points} \quad (20)$$

As the statistical models are trained to predict number of bike trips, which is a numeric and discrete variable, the models used in this study are regressors. The models studied are 1) multiple linear regression, a simple model that assigns coefficients to each predictors; 2) k-nearest neighbors regression, a non-parametric model that makes predictions based on similar observations in the train set; 3) random forest regression, which builds many uncorrelated decision trees and aggregates their results to form robust predictions; and 4) Adaptive Boosting (AdaBoost) regression, a boosting algorithm that iteratively builds a strong predictive model by learning from the errors of weaker models.

To evaluate the relative importance of predictors, the permutation importance of each predictor is calculated. The permutation importance is calculated by randomly shuffling the values of a single predictor among the data points and measuring the increase in mean-squared-error as values for that predictor can no longer contribute meaningfully to the prediction.

For the multiple linear regression and k-nearest neighbors models, the ‘Route’, ‘Year’, and ‘Hour’ predictors are treated as one-hot encoded categorical variables. ‘Weekend’ and ‘Special Date’ are binary, while ‘Precipitation’ and ‘UTCI’ are continuous. The random forest regression and AdaBoost regression used the same predictors, except ‘Hour’ which is treated as a discrete numeric variable.

## 3. Results

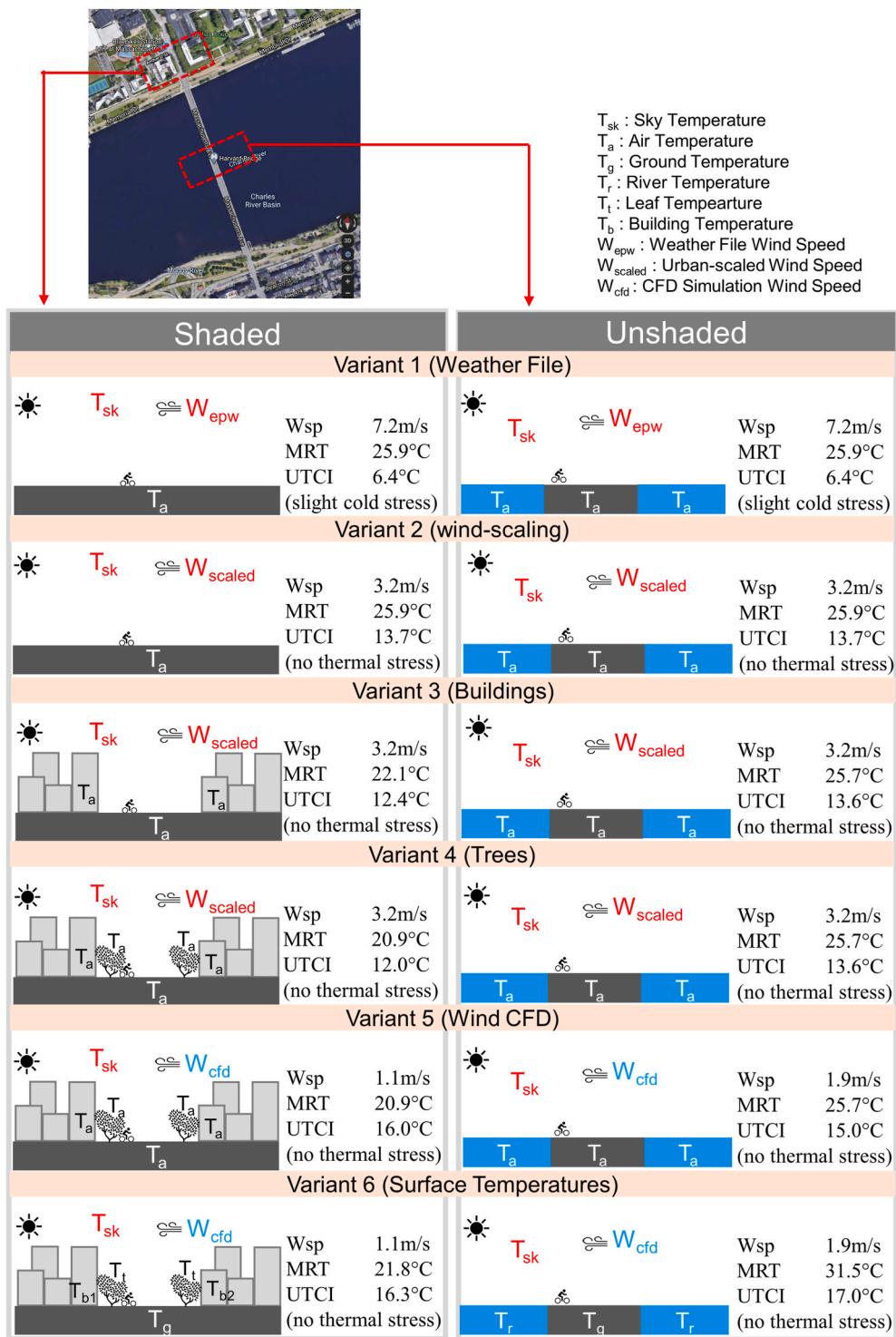
### 3.1. Overview

The six variants of UTCI simulations are run for the three bike routes over a span of 3 years. With Route 1 comprising the most diverse types of environments, such as an urban area with buildings, an exposed area over a bridge, and areas with trees, the results for Route 1 are closely analyzed and presented in this section. Furthermore, while three years of simulations was run to ensure a sufficiently large set of results were available for the statistical model in following sections, only results from the year 2019 are presented in detail in this section.

The six variants of UTCI results are compared on three scales: 1) singular probing points at a shaded and an unshaded region of the route at a specified point in time, 2) spatially along the whole route at a specified point in time, and 3) singular probing points at a shaded and an unshaded region over a whole year. Subsequently, metrics for aggregating UTCI along a route are introduced to prepare the UTCI results for the bike ridership study in the next chapter.

### 3.2. Point analysis

UTCI simulation results at 1pm on May 10, 2019, are compared at two points along Route 1, with one being shaded under buildings and trees, and the other exposed to sun and wind (Fig. 6). At this hour, the weather station at Boston Logan International Airport reported an air temperature of 14.4 °C and relative humidity of 72%. Results from Fig. 6 show that scaling down the wind-speed in Variant 2 significantly affects UTCI for both the shaded and unshaded points and causes the



**Fig. 6.** Illustrations of simulation assumptions for Variant 1 to Variant 6 and their effects on UTCI values at a shaded point and an unshaded point along Route 1 (year 2019).

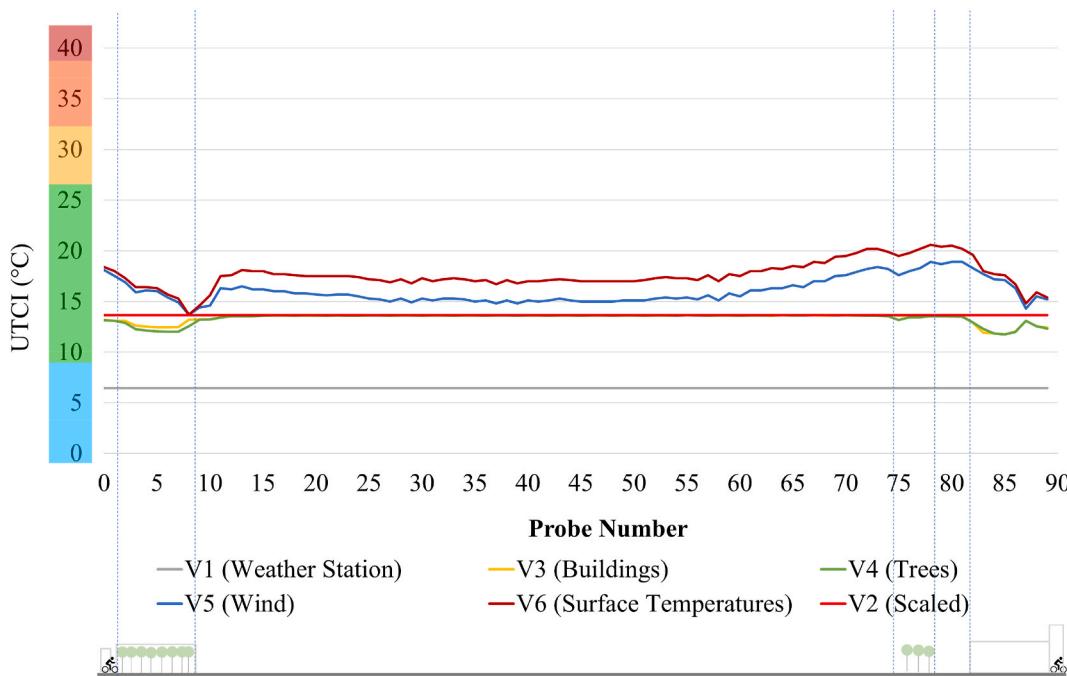
UTCIs to move up a thermal band from slight cold stress to no thermal stress. Simulating shading from buildings and trees affect the shaded point by lowering UTCI by a total of 1.7 °C, while no significant change was observed for the unshaded point. When CFD simulations were introduced in Variant 5, the change in UTCI from the previous variant was more significant for the shaded point than the unshaded. The effect of simulating surface temperatures in Variant 6 at this hour was more significant for the unshaded point, which observed a 2 °C increase in UTCI. For this particular point shown, the UTCI results between both

environments differed minimally, especially in Variant 5, as the wind speed during this hour was relatively low.

(Imagery ©2021 Google, Imagery ©2021 CNES/Airbus, MassGIS, Commonwealth of Massachusetts EOEA, Maxar Technologies, Sanborn, USDA Farm Service Agency, Map data ©2021 Google).

### 3.3. Route analysis

For the same time period, UTCI results along Route 1 are plotted in



**Fig. 7.** Simulated UTCI results along route 1 on May 10th 2019 at 11am for Variant 1 to Variant 6.

**Fig. 7.** Probing points are placed 10 m apart along the route, resulting in 90 points along route 1 that are numbered 0 through 89. Probes 0 to 9 are within a dense urban area at MIT's main campus, probes 10–80 are exposed on the Harvard bridge connecting Boston and Cambridge (with the exception of a few trees around probe 75), and probe 81 to 89 are within an urban area at Boston's Back Bay neighborhood.

Results in Fig. 7 reflect a similar pattern to the results in the point-based analysis in Section 2.3.1. Variant 1's UTCI, being calculated purely from the weather file and independent of geometry, is constant along the route at 6.4 °C. Variant 2's UTCI is uniformly scaled and is thus similarly homogeneous. Changes in UTCI along the route is observed from Variant 3 onwards, where geometry is accounted for. Where shading from buildings and trees are present (probe 0 to 9 and 75 to 89), the UTCI decreases. The effect of localized wind speeds is reflected in Variant 5, where UTCI increases to different extents at different points along the route. Considering surface temperatures in Variant 5 in this case causes UTCI to increase along the bridge to a larger extent than in the urban areas as the unshaded building/ground temperature on the exposed bridge deviates more than that in the shaded urban area.

The effects of increasing spatial resolution through each variant differs from hour to hour depending on the weather attributes. To understand the impact of different levels of spatial resolution on UTCI results, UTCI values are further analyzed on an annual scale in the next section.

### 3.4. Annual analysis

Hourly UTCI results over a whole year are analyzed at two along Route 1 (the same two points-of-interest as the point analysis). Fig. 8 presents annual heatmaps comparing the UTCI results for both points for all six variants. Annual hourly UTCI results for Variant 1 are categorized according to the UTCI assessment scale (Fig. 1) and plotted on a heatmap. Thereafter, differences in comfort levels between consecutive variants are plotted. An increase in heat stress, which is a decrease in cold stress, is plotted in orange. An increase in cold stress, which is also a decrease in heat stress, is plotted in blue.

The annual heatmap of Variant 1 UTCI reflects Boston's climate where one would experience cold stresses for most of January to May and November to December. The summer months between June and

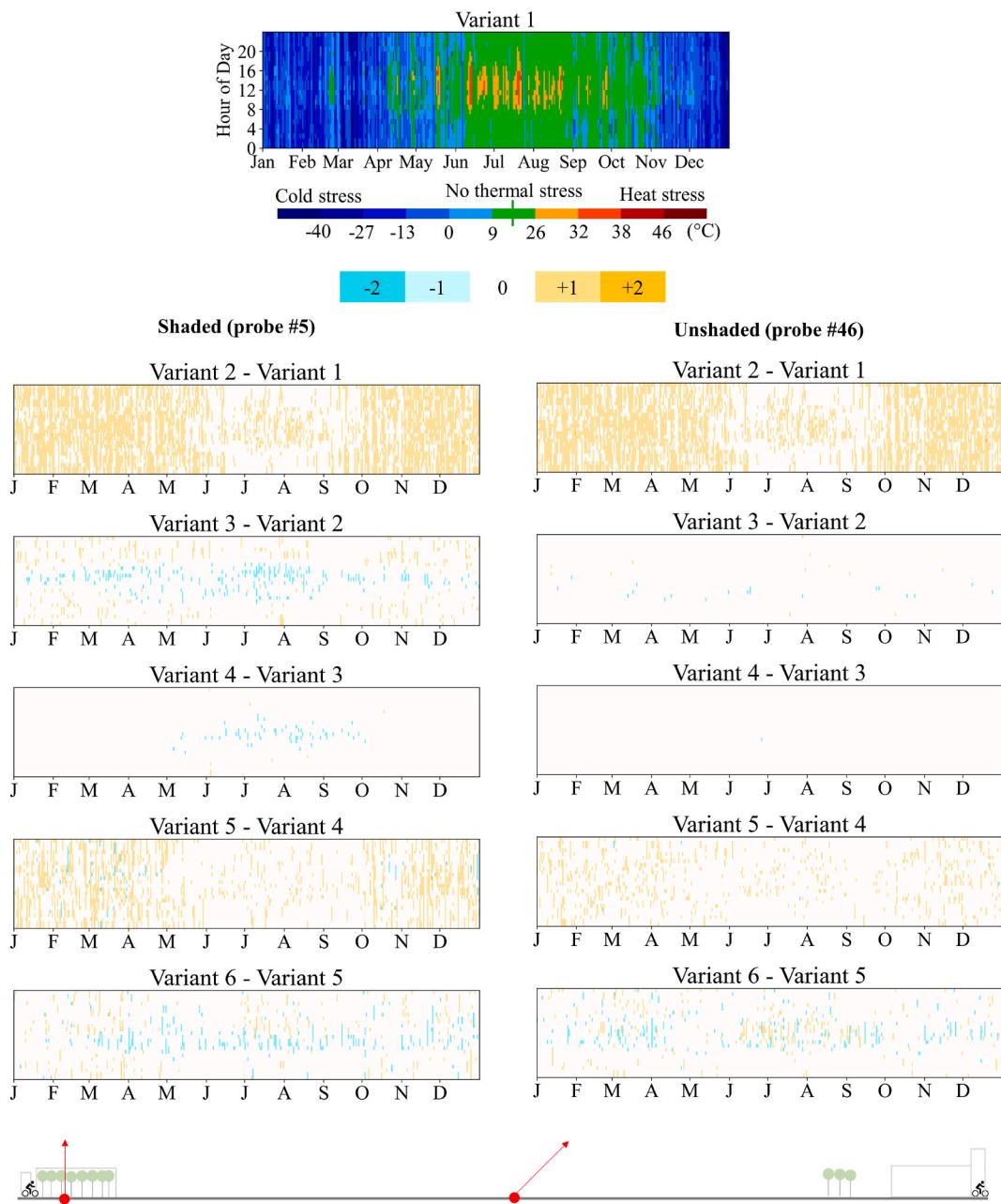
October shows large regions of no thermal stresses especially in the early mornings and late evening. Heat stresses are prevalent during the day in the summer months, especially in July and August.

The scaling of wind speeds to suit an urban area cause significant increase in heat stress throughout the year, as observed from the difference plot between Variant 1 and 2. This worsens heat stress in the summer when it is already warm, but improves comfort in the winter when it is cold. The difference plots between Variant 2 and 3 show that accounting for shade from buildings causes UTCI to be colder in the day when buildings reduce exposure to solar radiation and warmer in the nights since there is less exposure to the cool night sky. This decrease in UTCI in the day is beneficial in the summer as it reduces heat stress, but detrimental in the winter when it further worsens cold stress. Accounting for trees in Variant 4 causes comfort levels to shift towards colder bands when the trees are present at the shaded point and in season in the summer months.

The increase in UTCI due to Variant 5's wind simulation, as observed in the point analysis and route analysis, is also observed on an annual scale. The difference plot between Variant 4 and Variant 5 show that accounting for local wind speeds through CFD increases heat stress to a greater extent at the shaded area than the exposed area. This increase in heat stress observed is similar to the effect of applying wind-scaling in Variant 2, although to a lesser extent.

From Variant 5 to Variant 6, simulating surface temperatures along the route causes comfort levels to get warmer or colder at different times. From 8am to 12pm, when the sun is not shining overhead, the shaded point observes mostly increases in cold stress. However, after 12pm, during hours when the sun is at its highest point, the mean radiant temperature increases and contributes to greater heat stress. This pattern is not observed in the unshaded point, and the impact of simulating surface temperatures fluctuates between increasing heat stresses and cold stresses depending on the road surface temperatures.

Overall, the most significant differences in hourly comfort levels observed are from the inclusion of urban-scaled wind speeds in Variant 2 and Variant 5 (Fig. 9). For the shaded point, simulating wind speeds with CFD in Variant 5 still cause a 3.3% increase in 'no thermal stress' hours, despite the initial scaling implemented in Variant 2. However, for the unshaded point, the wind-scaling implemented in Variant 2 to 4 provided a reasonably good estimate of wind as the CFD simulation only



**Fig. 8.** Annual heatmaps (hour of day against day of year) of hourly UTCI and hourly differences in UTCI (as indicated above each heatmap) for two points along route 1 (a shaded point and an unshaded point).

contributed to a 1.3% increase in ‘no thermal stress’ hours. The next most significant differences in comfort levels was caused by the incorporation of surface temperatures in Variant 6 and this was observed for both the shaded and unshaded point. Accounting for shading from buildings and trees caused approximately 2% increase in comfortable hours in the shaded point, but caused no change for the unshaded point.

### 3.5. Relating bike ridership to thermal comfort

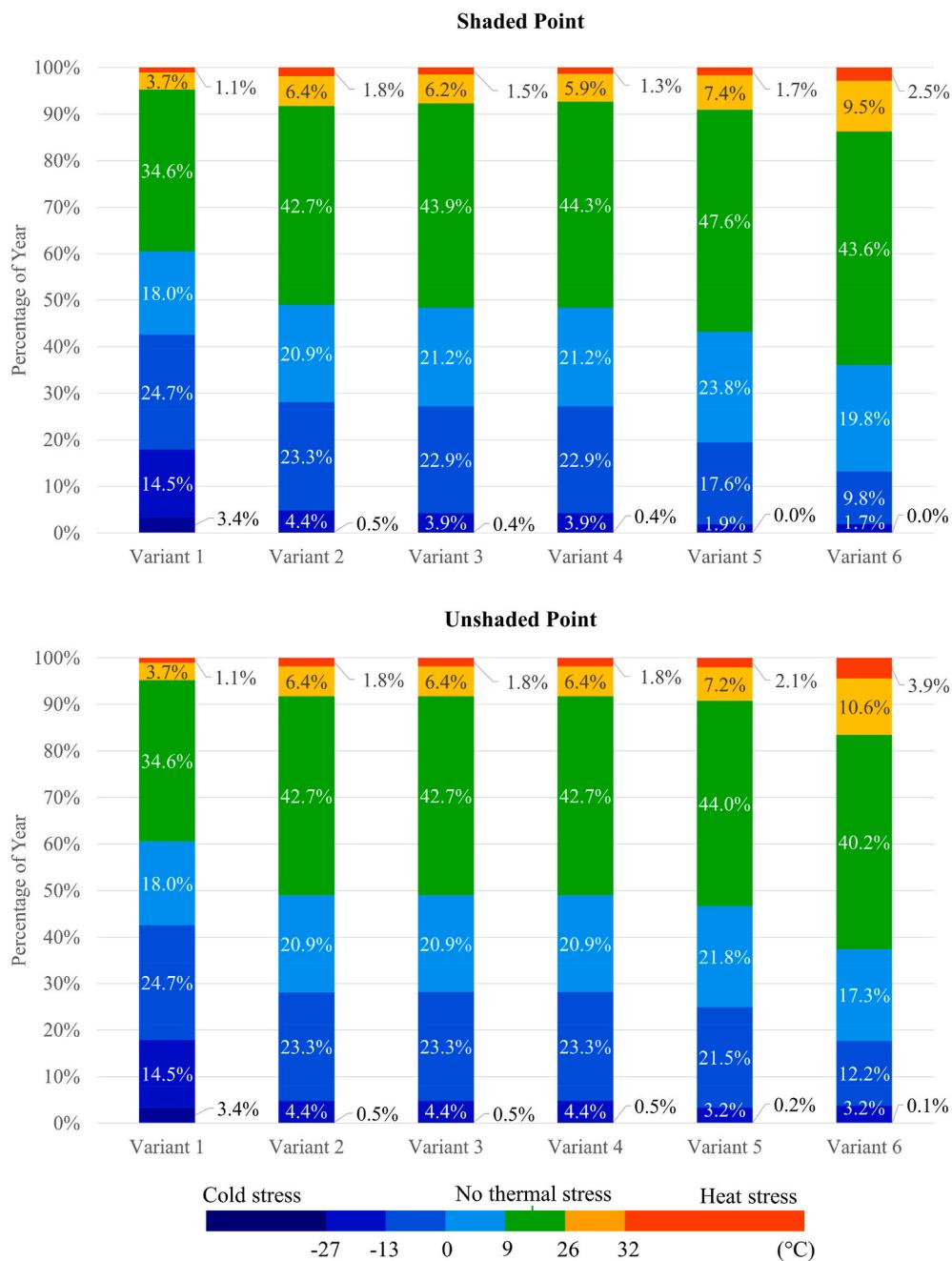
The Bluebikes dataset and UTCI simulation results were used to compute the average number of trips per hour against UTCI (Fig. 10). More trips occur during warmer hours, which suggests that cyclists are more likely to bike as cold stress decreases and heat stress increases.

Other predictors such as the demand and weather predictors affect bike ridership as well and should be accounted for when studying the relationship between average trips per hour and UTCI comfort levels.

Four key predictors are considered: hour of the day, route, special date, weekend, and precipitation.

Average trips per hour is plotted against hour of the day (i.e. 12am, 1am, 2am ... 11pm) for each route in both directions (Fig. 11). The six bar graphs show the bike ridership differs from hour to hour and between different routes. Routes 1a, 2a, and 3a show morning peaks when compared to routes 1b, 2b, and 3b. Routes 1a and 3a show peaks in demand in the morning when MIT students and staff start their day. Similarly, Route 2a shows a peak in demand in the morning when office goers head towards the train station at Davis Square.

With varying demand for bikes in each route, the relationship between UTCI and trip data is analyzed separately for each route. The following analysis focuses on Route 1b: MIT to Beacon St, and ensures that patterns observed between UTCI and trip data are not influenced by the route-type. Additionally, to account for variability in demand due to the time of the day, data within the 2pm–6pm time interval will be used



**Fig. 9.** Frequency of occurrence of each UTCI comfort level (percentage rounded to nearest whole) for each variant over a whole year (2019) at two points along route 1. High heat stress and high cold stress categories are (separately) combined due to low proportions.

to reduce variability.

The next predictors consider factors that affect the type of day that each hour falls on. The average trips per hour in Route 1a and 1b are plotted against the hour of day for school days and non-school days (including holidays and weekends) separately (Fig. 12). For both routes, more trips are taken during school days than non-school days, especially in the late afternoon hours.

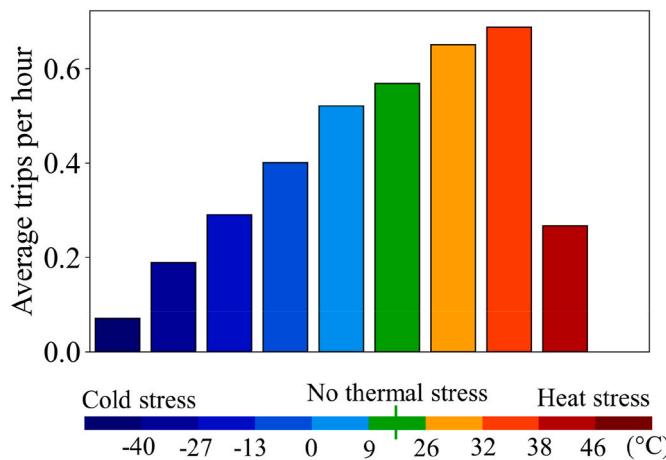
Lastly, the effect of precipitation is considered (Fig. 13) and the average number of trips per hour is lower with precipitation than without.

To rule out the effects of the aforementioned factors on average trips per hour in this study, the dataset is filtered for trips that are along Route 1b, occurring during MIT school days, on weekdays, and when there is no precipitation. The average trips per hour for each UTCI comfort level in this limited dataset is plotted in Fig. 14. The average number of trips

per hour peaks in the ‘no thermal stress’ comfort level is the highest and tapers down as cold stress and heat stress increase. The average number of trips per hour for the strong heat stress category (in red) does not conform to the trend, and is likely caused by outliers in the few data points falling into that category. With the exception of the outlier, this trend indicates that the more comfortable a person feels, as indicated by the UTCI comfort levels, the more likely they are to bike.

To compare between the different variants of UTCI, the same limited trip dataset is used to plot average trips per hour against the various variants of UTCI (Fig. 15). To fit the data to a parabolic regression, a less discrete form of UTCI bins is used. UTCI is binned into 3-degree bins (e.g. 0 °C–3 °C, 3 °C–6 °C, 6 °C–9 °C) and the average number of trips per hour is calculated for each bin.

As the number of data points within each UTCI bin is different, the quadratic polynomial is fit using weighted least squares regression. The



**Fig. 10.** Average number of trips per hour along all bike routes for each UTCI comfort band.

weight used for each UTCI bin is the number of data points contributing to the bin, allowing bins which have more data points to contribute more to the fit, and bins with less data points to contribute less. The Weighted  $R^2$  for all variants and all route UTCI metrics (start UTCI, worst UTCI, and worst segment UTCI) are relatively similar and range between 0.63 and 0.83. The best performing variant is Variant 5 average UTCI, as highlighted in red in Fig. 15. Variant 5 average UTCI accounts for local

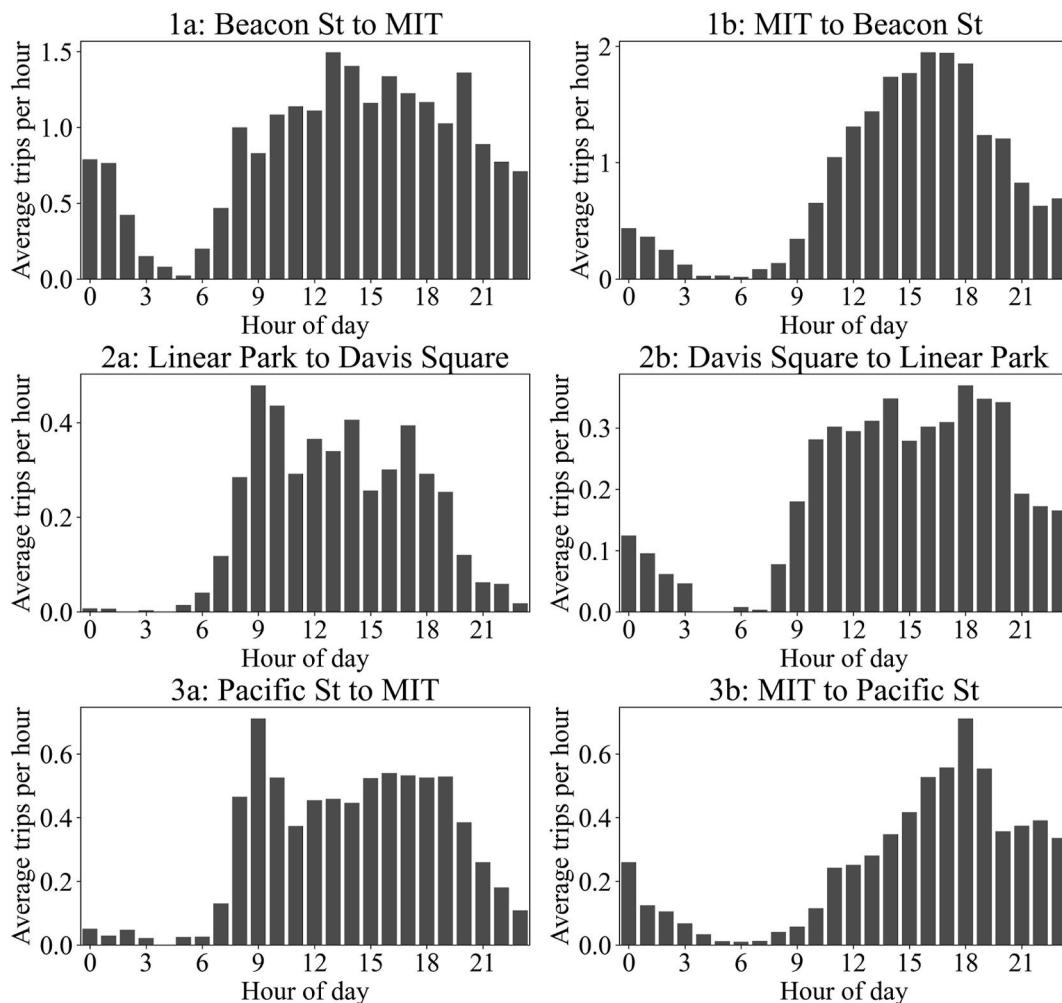
wind speeds in the UTCI simulations and averages all UTCI values along a route to obtain an effective route UTCI.

### 3.6. Statistical modeling

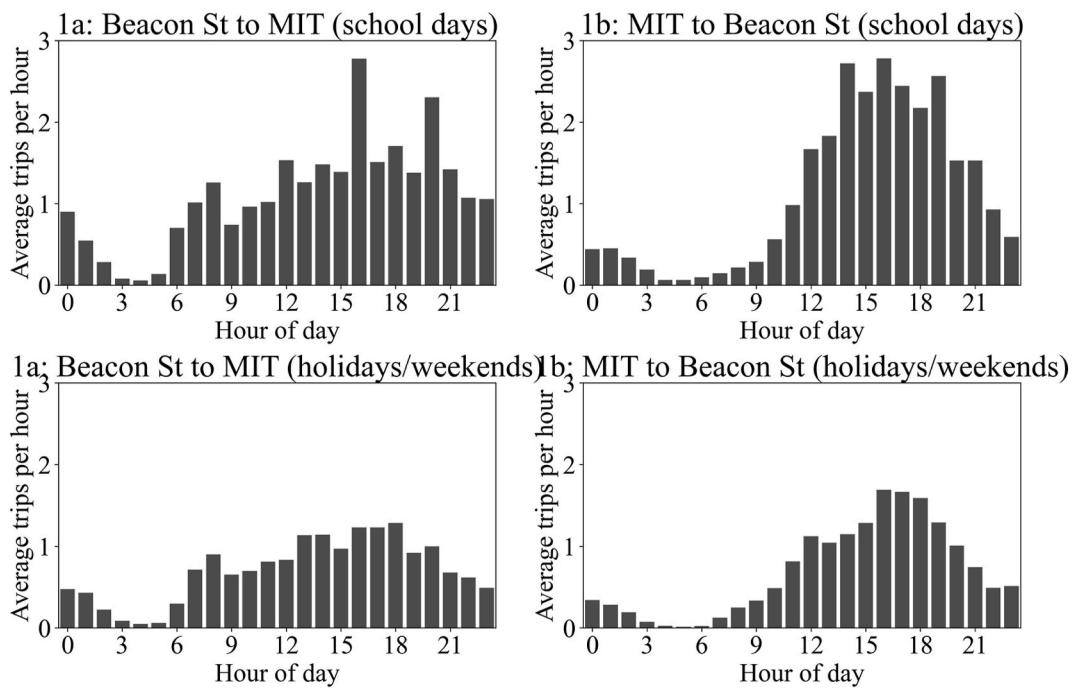
A statistical model is trained to predict the number of expected bike riders along routes for a given hour using various predictors that affect bike ridership, including UTCI. Previously, the relationship between UTCI and average trips per hour was explored after narrowing the dataset down to a specific set of conditions. While the approach taken is useful in understanding trip patterns within a narrow set of conditions, predicting ridership for a variety of predictor values is tedious with that approach. Training a statistical model to predict the number of bike trips for various combinations of predictors is useful for applications such as demand predictions to facilitate bike distributions in a bike sharing program.

The train RMSE and test RMSE results from all models are presented in Table 6. For each model, the different variants of UTCI performed almost identically, with a difference in RMSE of less than 0.01. The RMSE values for Variant 5 Average UTCI are presented. All models performed better than baseline, and the random forest regression performed best by a slight margin.

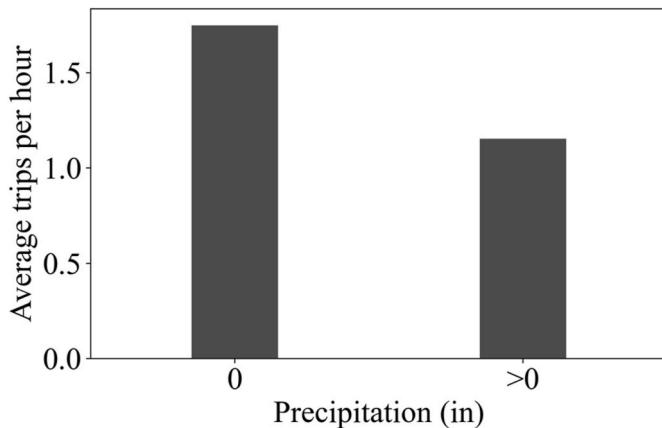
As the statistical model are regressors, the predictions by the models are continuous in nature and are not necessarily whole numbers. To study the difference between the actual trip counts and the predicted trip counts, the random forest model's predictions on the test set are rounded to the nearest integer. These rounded predictions are compared



**Fig. 11.** Average number of trips taken for each hour of the day along all routes, calculated from Bluebikes data from 2017 to 2019.



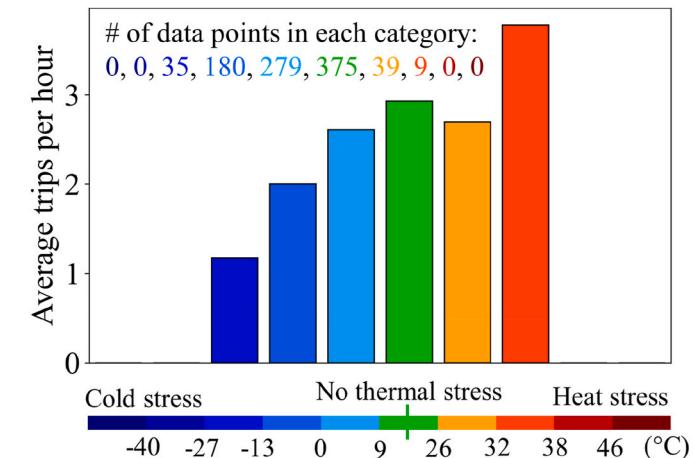
**Fig. 12.** Average trips along two routes (both directions) at MIT for MIT school days and MIT holidays/weekends. Results calculated from 2017 to 2019 data.



**Fig. 13.** Average number of trips per hour for all routes from 2017 to 2019 during hours with or without precipitation.

against the corresponding actual trip counts in a confusion matrix in Table 7. Each entry contains a count of the data points for which the random forest model outputs a certain predicted trip count for a given actual trip count. The counts are also given as a percentage of the size of the full test set. 47% of the test set is predicted accurately (i.e. 0 trips are predicted as 0 trips, 1 trip is predicted as 1 trip) and 42% of the test set is predicted with 1 trip less or more than the actual count. 75% of the dataset is predicted to have trip counts that are equal to or more than the actual trip count.

The importance of each predictor in the random forest regression was compared using the permutation importance (Table 8), which is a measure of the increase in the model's mean squared error (MSE) when each respective predictor is randomly shuffled. More important predictors would have higher permutation importance values. As the models with different variants of UTCI performed equally well in terms of RMSE, the permutation importance values are similarly equal (or very close) across variants. The permutation importance values for Variant 5 Average UTCI is shown in Table 8. 'Route' is by far the most important



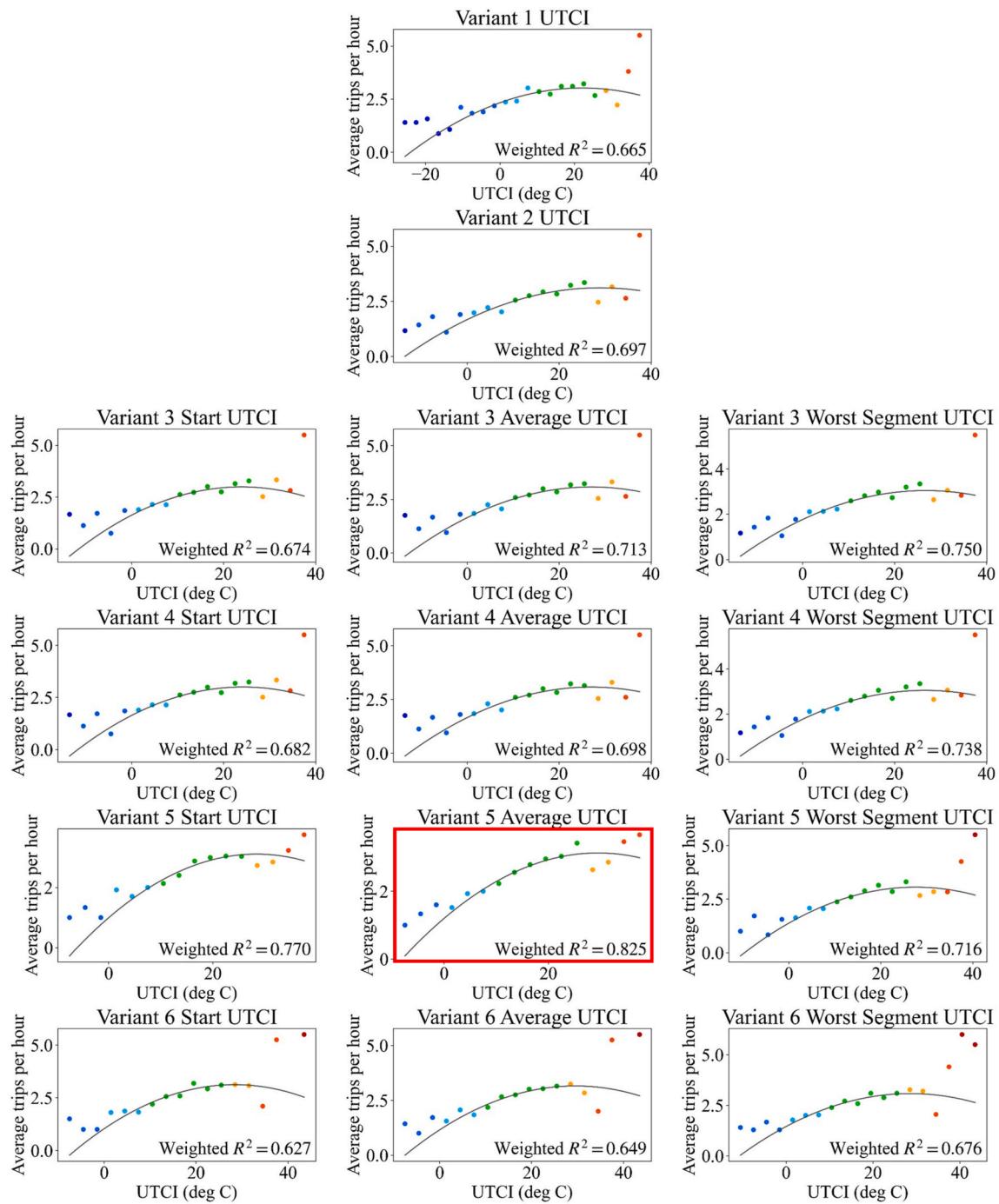
**Fig. 14.** Average number of trips per hour along route 1 for each comfort level, on MIT school days, in the afternoon 2–6pm and when there is no precipitation. The number of data points that fall within each comfort band (used to calculate the average trips per hour) is provided.

predictor, followed by 'special date' and 'hour'. These findings correspond to those found in the earlier analysis relating bike ridership to thermal comfort. UTCI comes fourth and is more impactful than 'weekend', 'year', and 'precipitation'. Surprisingly, 'Precipitation' is not an important predictor and is likely rarely used by the random forest regressor's decision trees.

#### 4. Discussion

The UTCI simulations show that on an hourly basis, different levels of spatial resolution do impact UTCI values. Depending on the type of space, whether shaded or unshaded, the variants have different impacts on hourly UTCI.

For a shaded portion of a bike route, considering shading from buildings and trees would have significant impact. The impact of shading can cause the UTCI to shift a comfort band up (warmer) or down



**Fig. 15.** Plots of average trips per hour against different variants of UTCI (in bins of 3 °C) for trips along Route 1 on MIT school days, in the afternoon between 2 and 6pm and when there is no precipitation.

**Table 6**  
RMSE results from baseline and statistical models.

Model	Train RMSE (# trips)	Test RMSE (# trips)
Baseline	1.52	1.51
Multi-linear Regression	1.13	1.12
K-Nearest Neighbors	1.02	1.06
Random Forest Regression	0.98	1.06
AdaBoost Regression	1.06	1.07

(colder) as seen in the annual analysis in Fig. 8. The impact of Variant 3, which considers shading from trees, is more significant in routes that are populated with trees such as Route 2, and only in warmer summer months when trees are in season. Overall, when the changes to hourly UTCI are considered on an annual basis, the increase in heat stress hours and increase in cold stress hours generally balances out. This causes the annual distribution of comfort levels from Variant 2 to Variant 4 to remain similar (Fig. 9).

Variant 2 and Variant 5, which incorporate a change in the method of calculating wind speed, show significant changes in UTCI. From Variant 1 to Variant 2, there is a large change in the proportion of comfortable hours of approximately 8%. From Variant 4 to Variant 5, the change is

**Table 7**

Confusion matrix of prediction accuracies on the test set for the random forest regression model.

Predicted Trip Count		0	1	2	3	4	5	6
Actual Trip Count	0	1312 (32%)	870 (21%)	94 (2%)	12 (0%)	5 (0%)	0 (0%)	0 (0%)
	1	423 (10%)	511 (12%)	90 (2%)	32 (1%)	2 (0%)	0 (0%)	0 (0%)
	2	71 (2%)	221 (5%)	90 (2%)	48 (1%)	4 (0%)	0 (0%)	0 (0%)
	3	20 (0%)	77 (2%)	54 (1%)	36 (1%)	0 (0%)	0 (0%)	0 (0%)
	4	4 (0%)	26 (1%)	35 (1%)	29 (1%)	2 (0%)	1 (0%)	0 (0%)
	5	0 (0%)	7 (0%)	16 (0%)	23 (1%)	2 (0%)	0 (0%)	0 (0%)
	≥ 6	0 (0%)	12 (0%)	13 (0%)	16 (0%)	2 (0%)	0 (0%)	1 (0%)

**Table 8**

Permutation importance of predictors in Random Forest Regression.

Predictor	Permutation Importance (increase in MSE)
Route	0.565
Special Date	0.209
Hour	0.116
UTCI (Variant 5 Average)	0.068
Weekend	0.051
Year	0.020
Precipitation	0.000

smaller but still noticeable at approximately 3% for the shaded point and 1% for the unshaded point. However, the scaling in Variant 2 does not fully account for the surrounding geometry especially in the shaded point which is surrounded by complex geometry. In such cases, running a CFD simulation would provide significant additional benefit.

In this study, the CFD simulation assumed a roughness length of 2 m in all wind directions, which describes a dense city environment. This roughness length more accurately represented the dense regions on either side of the bridge than the exposed region on the bridge. The simulated wind and UTCI results for Variant 5 could possibly come close to reflecting actual conditions in the shaded area but may not be as accurate for the exposed area. This uniformity in roughness length explains why the urban-scaled UTCI in Variant 2 was similar to Variant 5 along the bridge despite its exposed nature. To further increase the accuracy of the CFD simulation, one could consider introducing different roughness length assumptions for each simulated wind direction. The results of the study quantitatively show that scaling the wind speeds to reflect the surrounding environment partially substitutes the need to run a full CFD simulation, which many may find to be too complex for their needs. However, if a designer intends to study the effect of urban interventions such as local wind shields, a full CFD simulation will be necessary.

Variant 6, which simulates surface temperatures, impacts hourly UTCI as shown in the route-based plot (Fig. 7). The annual analysis also shows a significant number of hours shifting a comfort band up or down (Fig. 8). The direction in which Variant 6 causes UTCI to shift towards largely depends on the building temperature. Generally, when Variant 6 is warmer than Variant 5, the aggregate surface temperature of buildings and the ground is higher than air temperature. When Variant 6 is colder than Variant 5, the reverse is true. This correlation is observed for 90% of all hours for Route 1 in 2019. On an annual basis, there is a significant change in comfort hours and in the case of Route 1 in 2019, a 4% increase in comfortable hours was observed for both shaded and unshaded points.

Overall, the results of the UTCI simulations show that considering contextual shading and surface temperatures in the MRT calculation, as well as accounting for local wind speeds, is necessary to accurately predict UTCI on an hourly basis, especially in urban areas. However, if one were only concerned about the distribution of comfort levels annually, considering wind speeds, whether through scaling in Variant 2 or CFD in Variant 5, causes the greatest impact. CFD simulations may lead to accurate and localized UTCI results, but preparing the geometric

model and running the simulations require a considerable amount of effort. The amount of data that has to be managed can also be daunting. A scaling factor (as used in Variant 2) is useful in cutting down on simulation time while ensuring that UTCI results are relatively close to actual thermal comfort conditions.

The study between UTCI results and bike ridership along Route 1 during peak afternoon hours on MIT school days show that the average number of trips per hour increases with thermal comfort. The more comfortable a route is, the more trips are taken on average (Fig. 15). This reinforces the importance of thermal comfort in encouraging human-powered transport in cities.

While most variants of UTCI fit similarly well to the parabolic regression in Fig. 15, Variant 5 Average UTCI performed best with a  $R^2$  value of 0.83. The relatively stronger relationship between Variant 5 UTCI and bike ridership agrees with UTCI simulation findings on the importance of considering local wind speeds in calculating UTCI in urban spaces. The less significant differences in  $R^2$  values among the other variants is likely because the UTCI simulation results change in a similar manner across a year when the physical accuracy is increased. For example, when urban-scaled wind speeds are introduced from Variant 1 to Variant 2, the UTCI values undergo a ‘translation’ in value as they increase across the year. The statistical methods used in this study are able to adapt to such changes – although the actual values of UTCI change, the difference in UTCI between two particular hours is used by the models to make predictions and the differences experience much smaller changes between UTCI variants.

In attempting to predict bike ridership using statistical models, the models that were explored performed better than the baseline model. The test RMSE of the best performing model, a random forest regressor, is 1.06 trips (Table 6). When the accuracy of predictions is computed, 47% of the hourly trip counts are predicted correctly (Table 7). While the accuracy of the best model is relatively low, it is expected as the bike ridership data did not come with any information about individual cyclists that would be useful in predicting ridership, other than their subscriber status which this study filtered for. The model assumes that cyclists along each route have the same routines and travel patterns when in reality they likely do not. Information about the cyclists’ routines is particularly important in the case of a bike route that is frequented by students whose attendance strongly varies over the course of the year as well as over several years. Hence, being able to predict trips correctly half the time without insight into the rider can be considered relatively good.

For the purposes of estimating demand for Bluebikes bicycles to ensure sufficient bikes are available at dock stations for customers, the proportion of the dataset that is conservatively predicted would be of interest. This would be when the predicted trip count is greater than or equal to the actual trip count. In this case, the model performs relatively well and conservatively predicts demand 75% of the time.

The influence that UTCI has on model predictions can be inferred from permutation importance values in Table 8. The three most important factors are the route of interest and the type (i.e. special date) and time of day. These are demand predictors that influences whether a person has a reason to make a trip. Once demand is established, UTCI

becomes a key factor in deciding whether a person chooses to bike. This supports the need to model and predict thermal comfort in urban planning and design to encourage people to bike or walk more. Precipitation, on the other hand, is found to be an unimportant predictor, which is unexpected as people are likely to be deterred from biking when it is raining.

The unidentifiable nature of the Bluebikes trip data presents significant limitations to this study, both in the accuracy of predicting bike ridership and in understanding the importance of UTCI in mobility mode choice. If anonymized ridership data were available, user behavior and patterns could be analyzed to study when and why frequent cyclists decide to switch travel modes. More work is needed to quantitatively relate UTCI to bike ridership or pedestrian movement patterns, and encrypted but identifiable trip data would provide the opportunity to develop a better-performing model that considers different cyclist archetypes. User behavior and patterns could also be analyzed to study when and why frequent cyclists decide to switch travel modes.

## 5. Conclusion

This study explored six variants of UTCI simulations with increasing levels of physical accuracy and spatial resolution. The variants start with a calculation that is purely from a local weather file, then includes scaling of wind speeds based on the surrounding terrain type. Subsequently, spatial resolution is increased by considering shading from geometry such as buildings and trees. Next, local wind speeds are calculated using CFD simulations, and surrounding surface temperatures are simulated. The UTCI results are then used to relate spatial-temporal thermal comfort to bike ridership in the Boston area through the Bluebikes public bike sharing program.

This study demonstrates the importance of spatial refinement in simulating UTCI. At the very least, designers should scale wind speeds according to the type of terrain. If one were considering dense urban areas, a more intensive CFD simulation would improve accuracy. However, this study is limited to the Boston area and further work is required to understand the impacts of the various refinement levels in different climates. It motivates future research into efficient simulation methods or rules-of-thumb for deriving spatial-temporal UTCI values. Incorporating UTCI into prediction models for bike ridership improves prediction accuracy, and future work into building a more accurate model would motivate planners and policymakers to design thermally comfortable environments that encourage human-powered transportation.

## CRediT authorship contribution statement

**Elizabeth Young:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Patrick Kastner:** Methodology, Software, Writing – review & editing. **Timur Dogan:** Methodology, Supervision, Writing – review & editing. **Ata Chokhachian:** Methodology, Software, Writing – review & editing. **Sarah Mokhtar:** Conceptualization, Methodology, Writing – review & editing. **Christoph Reinhart:** Conceptualization, Funding acquisition, Methodology, Software, Supervision, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The authors thank Sarah King and Cameron Reaves for their

involvement in the statistical modeling as part of the Harvard University CS109a Introduction to Data Science class, and Dominick Tribone of Lyft/Bluebikes for providing the Bluebikes availability data. This work was supported by the MIT Portugal Program [C-Tech 6942782].

## References

- [1] United Nations, 2018 revision of world urbanization prospects [Online]. Available: <https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html>, 2018.
- [2] U.S. Environmental Protection Agency, Fast Facts: U.S. Transportation Sector Greenhouse Gas Emissions, 1990–2019 (EPA-420-F-21-049, June 2021), Jun. 2021.
- [3] R. Buehler, T. Götschi, M. Winters, Moving toward active transportation: how policies can encourage walking and bicycling Promoting activity-friendly communities. Moving toward Active Transportation: how Policies Can Encourage Walking and Bicycling, Act. Living Res. (2016), <https://doi.org/10.5167/uzh-128504>.
- [4] R. Ewing, et al., Growing Cooler: the Evidence on Urban Development and Climate Change, 2007.
- [5] L. Bain, B. Gray, D. Rodgers, Living Streets : Strategies for Crafting Public Space, John Wiley & Sons, 2012.
- [6] U.S. Department of Transportation, Completed Streets. <https://www.pedbikeinfo.org/topics/completestreets.cfm>, 2021. (Accessed 28 April 2021).
- [7] CIVITAS, Car-independent lifestyles. <https://civitas.eu/TG/car-independent-lifestyles>, 2021. (Accessed 28 April 2021).
- [8] T. Combs, Local Actions to Support Walking and Cycling during Social Distancing Dataset, 2020.
- [9] J. Gehl, Life between Buildings : Using Public Space, Island Press, 2011.
- [10] D.J. Wuebbles, et al., Executive Summary. Climate Science Special Report: Fourth National Climate Assessment, Volume I, 2017, <https://doi.org/10.7930/J0D5CTG>.
- [11] L. Chen, E. Ng, Outdoor thermal comfort and outdoor activities: a review of research in the past decade, Cities 29 (2) (2012) 118–125, <https://doi.org/10.1016/j.cities.2011.08.006>.
- [12] K. Blazejczyk, Y. Epstein, G. Jendritzky, H. Staiger, B. Tinz, Comparison of UTCI to selected thermal indices, Int. J. Biometeorol. 56 (3) (2012) 515–535, <https://doi.org/10.1007/s00484-011-0453-2>.
- [13] P. Bröde, UTCI documents. [http://www.utci.org/utci\\_doku.php](http://www.utci.org/utci_doku.php), Jul. 2014. (Accessed 28 April 2021).
- [14] C.F. Reinhart, J. Dhariwal, K. Gero, Biometeorological indices explain outside dwelling patterns based on Wi-Fi data in support of sustainable urban planning, Build. Environ. 126 (2017) 422–430, <https://doi.org/10.1016/j.buildenv.2017.10.026>.
- [15] J. Dhariwal, P. Manandhar, L. Bande, P. Marpu, P. Armstrong, C.F. Reinhart, Evaluating the effectiveness of outdoor evaporative cooling in a hot, arid climate, Build. Environ. 150 (2019) 281–288, <https://doi.org/10.1016/j.buildenv.2019.01.016>.
- [16] R. Paolini, A.G. Mainini, T. Poli, L. Vercesi, Assessment of thermal stress in a street canyon in pedestrian area with or without canopy shading,” in, Energy Procedia 48 (2014) 1570–1575, <https://doi.org/10.1016/j.egypro.2014.02.177>.
- [17] L.L.C. Solemma, DIVA, Apr. 28, <https://www.solemma.com/diva>, 2021.
- [18] A. Chokhachian, M. Hiller, PANDO: Parametric Tool for Simulating Soil-Plant Atmosphere of Tree Canopies in Grasshopper, May 2020.
- [19] P. Kastner, T. Dogan, in: A cylindrical meshing methodology for annual urban computational fluid dynamics simulations 13, Jan. 2019, pp. 59–68, <https://doi.org/10.1080/19401493.2019.1692906>, 1.
- [20] T. Dogan, P. Kastner, R. Mermelstein, “Surfer, A fast simulation algorithm to predict surface temperatures and mean radiant temperatures in large urban models, Build. Environ. 196 (Jun. 2021) 107762, <https://doi.org/10.1016/j.buildenv.2021.107762>.
- [21] Bluebikes Boston, Bluebikes System data, Apr. 28, <https://www.bluebikes.com/system-data>.
- [22] L.K. Lawrie, D.B. Crawley, Development of global typical meteorological years (TMYx). <http://climate.onebuilding.org/contact/default.html>, 2019. (Accessed 11 May 2021).
- [23] W. Kessling, M. Engelhardt, D. Kiehlmann, The Human Bio-Meteorological Chart A Design Tool for Outdoor Thermal Comfort, 2013.
- [24] M.G. Lawrence, The relationship between relative humidity and the dewpoint temperature in moist air: a simple conversion and applications, Bull. Am. Meteorol. Soc. 86 (2) (Feb. 2005) 225–233, <https://doi.org/10.1175/BAMS-86-2-225>.
- [25] S.I. Tanabe, C. Narita, Y. Ozeki, M. Konishi, Effective radiation area of human body calculated by a numerical simulation, Energy Build. 32 (2) (Jul. 2000) 205–215, [https://doi.org/10.1016/S0378-7788\(00\)00045-1](https://doi.org/10.1016/S0378-7788(00)00045-1).
- [26] C.F. Reinhart, O. Walkenhorst, Validation of dynamic RADIANCE-based daylight simulations for a test office with external blinds, Energy Build. 33 (7) (2001) 683–697, [https://doi.org/10.1016/S0378-7788\(01\)00058-5](https://doi.org/10.1016/S0378-7788(01)00058-5).
- [27] Cambridge Department of Public Works, Cambridge tree walk, Feb. 07, <https://www.cambridgema.gov/Departments/PublicWorks/News/2013/09/streettrees-isomapping>, 2020. (Accessed 29 April 2021).
- [28] City of Somerville Urban Forestry Division, City of Somerville tree inventory, <https://www.somervillema.gov/departments/ospcd/psuf/urban-forestry>, 2018. (Accessed 29 April 2021).
- [29] Google, “Google maps, 2020.

- [30] C.Y. Park, et al., Variations in pedestrian mean radiant temperature based on the spacing and size of street trees, *Sustain. Cities Soc.* 48 (Jul. 2019) 101521, <https://doi.org/10.1016/J.SCS.2019.101521>.
- [31] The OpenFOAM Foundation Ltd, “OpenFOAM. <https://openfoam.org/>”, 2021. (Accessed 8 May 2021).
- [32] P. Bröde, et al., Deriving the operational procedure for the universal thermal climate Index (UTCI), *Int. J. Biometeorol.* 56 (3) (May 2012) 481–494, <https://doi.org/10.1007/s00484-011-0454-1>.
- [33] L.L.C. Solemma, ClimateStudio. <https://www.solemma.com/climatestudio>, 2021. (Accessed 11 May 2021).
- [34] E. Arens, T. Hoyt, X. Zhou, L. Huang, H. Zhang, S. Schiavon, Modeling the comfort effects of short-wave solar radiation indoors, *Build. Environ.* 88 (2015) 3–9, <https://doi.org/10.1016/j.buildenv.2014.09.004>, 6.
- [35] Massachusetts Institute of Technology, MIT bulletin archive. <http://catalog.mit.edu/archive/>, 2021. (Accessed 11 May 2021).