A generalizability analysis of a data-driven method for the Urban Heat Island phenomenon assessment

M. Pena Acosta[1], F. Vahdatikhaki[1], J. Santos[1], A. Hammad[2], A.G. Dorée[1]

[1]Department of Construction Management and Engineering, University of Twente, Enschede, the Netherlands
[2]Concordia Institute for Information Systems Engineering, Montreal, Quebec, Canada
E-mail:{m.penaacosta@utwente.nl, f.vahdatikhaki@utwente.nl, j.m.oliveiradossantos@utwente.nl, Amin.hammad@concordia.ca, a.g.doree@utwente.nl}

Abstract –

Cities worldwide are experiencing increasing temperatures due to the urban heat island (UHI) phenomenon. UHI is caused by the replacement of natural land surfaces with man-made dark surfaces. Among others, it causes a number of public health problems associated with heat events, particularly in the construction sector, where construction workers are more likely to die from heat-related illnesses compared to other industries. To address the negative effects of this phenomenon, researchers around the world have proposed different alternatives for studying the effects of UHI. Among these methods, data-driven approaches are becoming increasingly popular. However, as with all data-driven models, there is always the question of the extent to which they are generalizable. To answer this question, this research work applies the data-driven UHI assessment framework previously proposed by the authors to the cities of Montreal, Canada, and Apeldoorn, the Netherlands, in five different scenarios. The results showed that while the data-driven models have good prediction capabilities within the scope of the training data set, they do not demonstrate good generalizability on the testing data from a different context. Also, the results of this research highlighted that as cities continue to grow, there is an urgent need to standardize the understanding and assessment of the UHI at a pedestrian level. The intrinsic differences in how UHIs are assessed and tackled worldwide can create confusion about the phenomenon, and limits the applicability and generalizability of data-driven approaches.

Keywords –

Data-driven methods; decision trees; generalizability; urban heat maps; mitigation policies.

1 Introduction

The Urban Heat Island (UHI) phenomenon refers to the temperature difference between the outskirts and the inner city caused by the replacement of natural land surfaces with dark man-made surfaces [1]. As a result of these changes, cities are more prone to store solar radiation causing a series of public health problems associated with heat events [2]. The construction sector is no stranger to the harmful effects of this phenomenon. Construction workers, who often have to cope with high temperatures as part of their daily operations, are more likely to die from heat-related illness compared with other industries [3]. To further exacerbate the situation, there have been thirty-eight heat waves in Europe in the last century, seventeen of them in the last decade. The heat wave of 2003 caused 70,000 excess deaths over four months in Central and Western Europe [4]. Yet, cities are only projected to continue expanding. By 2050, cities are expected to shelter 68% of the world's population [5].

To counteract the negative effects of this phenomenon, researchers around the world have proposed different alternatives for the study of UHI effects. For example, physics-based simulation models try to mimic the thermal exchanges between urban surfaces and air temperatures. Yet, generating accurate and detailed simulations requires databases with three-dimensional representations of the built environment, resulting in very expensive simulations in terms of both computing power and time. Additionally, the complex interplay between the morphological characteristics of inner cities, socio-economic factors and UHI makes it difficult to develop an accurate and comprehensive physics-based model. Finally, the physics-based models are commonly too complex to be used by regular urban planners, making it difficult to incorporate UHI concerns into design and planning of cities.

In recent years, data-driven methods have gained popularity in solving complex multi-dimensional

problems. For example, Sutjaritvorakul et al. [6] developed a data-driven approach to detect workers from cameras installed on cranes. Similarly, Langroodi et al. [7] presented a data-driven approach to monitor the activities of construction equipment. Nutkiewicz et al. [8] developed a methodology to characterize and model the energy performance of buildings at multiple spatial and temporal scales. Regarding the domain of urban climates, a handful of machine-learning approaches have been developed in order to predict temperatures in specific urban settings [9-13]. For instance, Vulova et al. [10] developed a deep learning approach to discover hot spots in Berlin, Germany. The authors of this paper have also proposed a data-driven framework for the development of a user-friendly decision-support tool for UHI-sensitive urban planning that uses publicly available datasets [14]. Although this framework and similar ones developed by other researchers are shown to be effective and very promising, there is always the question of how generalizable these models are. This common question to all data-driven models is relevant because data-driven models are often trained and tested on the data collected from a limited geographical scope. Whether or not the intrinsic correlations and patterns discovered in the data are generalizable to other contexts is a question seldom addressed. Therefore, the aim of this paper is to assess the extent to which data-driven UHI assessment models can be generalized for different urban contexts. To do so, the data-driven UHI assessment framework previously proposed the authors [14], which relies exclusively on publicly available data, is applied to two different cities in five different scenarios. The studied cities are Montreal in Canada, and the Apeldoorn in the Netherlands. Apeldoorn is a medium-sized city, the 11th largest municipality in the Netherlands, with 150.000 inhabitants (2020). Located in the middle the country and the nature reserve De Hoge Veluwe, it presents a unique combination of greenery and built environment. In terms of infrastructure, the city center is densely built, having buildings with an average height ranging from 10 to 15 m. The construction materials of the building facades are homogeneous, varying from the dark to the light colors of traditional Dutch bricks. The city of Montreal, located in the southeast of Canada, is considered a metropolitan city, with a wide-ranging urban geometry. It covers an area of approximately 499 km^2, with a population density of 3.9 inhabitants/km^2.

Montreal and Apeldoorn constitute a good comparative basis because they are located in different climatic regions, have different scales in terms of size and population, and have distinctive urban morphologies and socio-economic characteristics. The remainder of the paper is structured as follows. First, in the interest of completeness, the data driven framework proposed by the authors is explained briefly. Then, the two case studies

are briefly explained. This is followed by the presentation of the results of the comparison of different scenarios. Finally, the conclusions and future work are presented.

2 Data-driven UHI assessment framework

As shown in Figure 1, the methodology consists of three main steps, namely: (1) data collection, (2) data preparation, and (3) model development. For the completeness of this research, the three steps are explained briefly in this section.

Figure 1: Overall framework [adapted from 14]

2.1 Data collection

The factors affecting UHI at a micro level (i.e., street level) can be categorized in three groups: (1) environmental factors, which have been described in the literature as uncontrollable factors [15], (2) socio-economic factors, and (3) urban morphology factors. Given that socio-economic and morphology factors have the greatest potential to be influenced by policies and urban planning decisions, they are the main target of the data collection strategy.

Table 2 presents the types of public data that are used for the development of the data-driven model. These data are used as the features (i.e., independent variables) of the data-driven model to assess the UHI effect at the street level. The sources from which these data were collected for the case of Apeldoorn are also presented in Table 2. In turn, the temperature variations in the city are used as labels (i.e., dependent variable) in the data-driven model and can be represented either in terms of the Physical Equivalent Temperature (PET) or Land Surface Temperature (LST). In the Netherlands, PET maps are widely used by the local authorities as a metric for the assessment of the heat stress.

2.2 Data preparation

Once the required data are collected, they first need to be processed to prepare them for the model development. To mine the urban features presented in Table 2, all features must be reduced to the street level. This requires the street jurisdiction to be determined first. This is the area surrounding the street segment from which all the urban features affecting the UHI can be assembled (e.g., building height, population, width of the

street, percentage of greenery, etc.). As shown in Figure 3, the street jurisdiction was based on the average width of the streets plus the minimum distance between the fence of each building and the adjacent street. For example, in the case of Apeldoorn, the street jurisdiction is defined by a buffer size of 15 m (the blue area in Figure 3 represents the buffer size).

Table 2. Primary features

Category	Primary features	Source
Socio-economic factors	Land use	pdok, land use [16]
	Building materials	Manual collection
	Population density	WorldPop Data, 2020 [17]
	Transportation (i.e., traffic flow)	Not available
Urban morphology	Building geometry	pdok, 3D building [18]
	Average building height to street width ratio (H/W)	NWB Wegen, pdok [19]
	Built-up density	Computation based on street jurisdiction
	Water bodies/cool sinks	
	Vegetation/green spaces	

Figure 2. Overview of the data preparation process [adapted from 14]

Figure 2 presents an overview of the data preparation

process. In a nutshell, all the socio-economic and urban morphology features were computed as follows: building, water, and vegetation density are calculated by taking the average overlapping area of each feature with the area of the corresponding street jurisdiction. Likewise, the population density and land distribution are calculated by projecting the administrative unit-based census data into geospatial gridded cell datasets and computing the average per street jurisdiction.

Regarding the data preparation of building materials, this is done by manually assigning a material type of each building in the city. This process can be conducted via Google Earth.
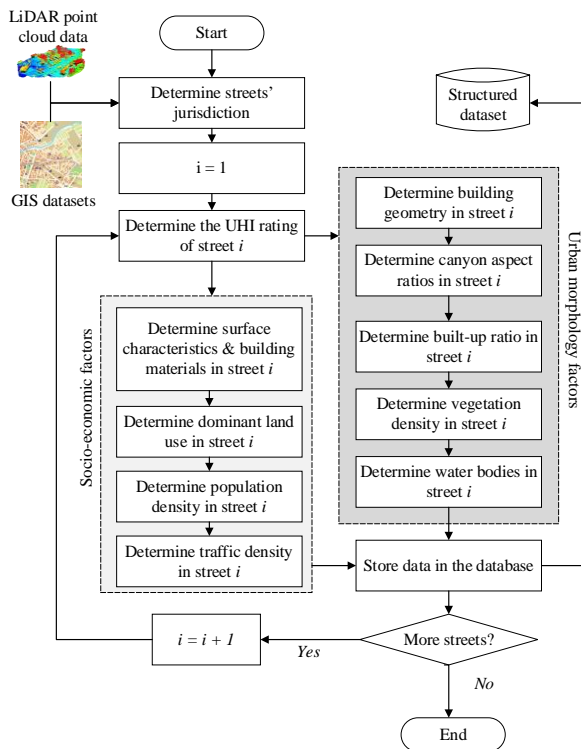

Figure 3. Average street section [images adapted from 20]

2.3 Model Development

In the model development, which is shown in Figure 4, the datasets are firstly split into training and testing sets to train the regressors. For this study, Decision Tree (DT) is used as the principle regressor. The hyperparameters of the DT model are set by default. This is followed by an optimization process in which the minimum number of levels of the tree (max-depth) is selected randomly, while minimizing the number of features that can successfully boost the learning process. Once the performance threshold is reached, the most important features, with the minimum max-depth, from the best performing models are selected.

3 Case studies

As mentioned earlier, the cities of Montreal and Apeldoorn were selected for the study that aims to assess the generalizability of data-driven models. The details of Montreal data were explained elaborately in the previous work of the authors [14]. For the case of Apeldoorn, the labels for this case study was retrieved from the work of Koopmans et al. [21].

It should be highlighted that label data of Montreal

was available in the form prescribed by the Jenks Natural Breaks Classification method [22]. In order to compare the two cities, PET temperature data of Apeldoorn were also classified using the Jenks Natural Classification Method. In a nutshell, this method splits the data in such a way that the mean standard deviation of data points in each category is minimized, while the standard deviation of distances of each data point to the means of the other categories is maximized. Figure 5 shows the classification of 2,768,169 data points into nine categories of PET temperatures.
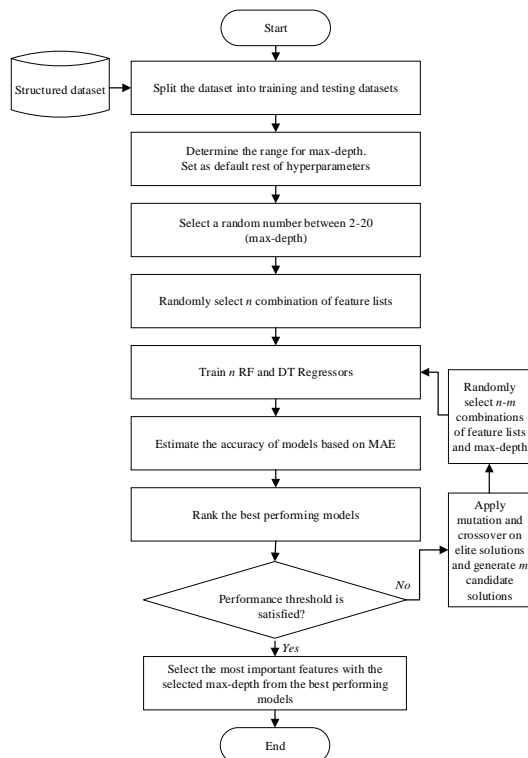

Figure 4. Overview of the model development [14]

Given the importance of the interpretability of the data-driven model in the context of urban climates [14], a DT regressor with a max-depth of 7 was implemented for all scenarios (to allow for the interpretability of each level of the DT). The other hyperparameters were initialized with the default configuration. The main primary features common to the two datasets were used to train the models and, subsequently, to compare their performance. These features are the following: building

façade materials, predominant land use per street jurisdiction, population count, average building height, vegetation, and building densities. Each dataset was divided into training (75%) and testing datasets (25%). Overall, 7,986 vector data instances were collected, of which 5,578 belong to Montreal and 2,408 belong to Apeldoorn.
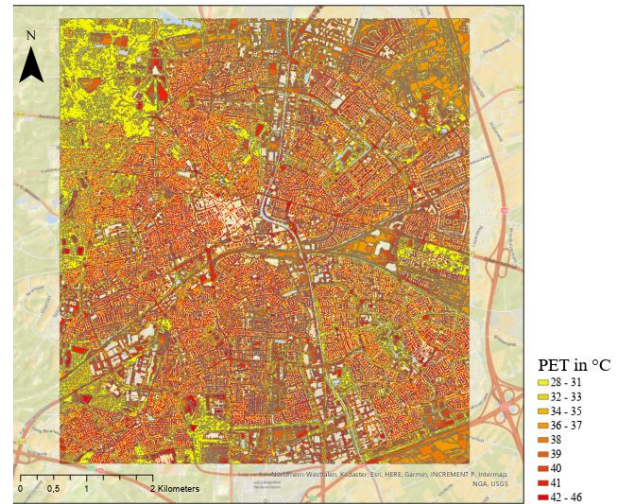


Figure 5. PET map of Apeldoorn

3.1 Comparison of the Models

Five different models are developed to analyse the scenarios presented in Table 1.

Table 1. Scenarios studied in this research

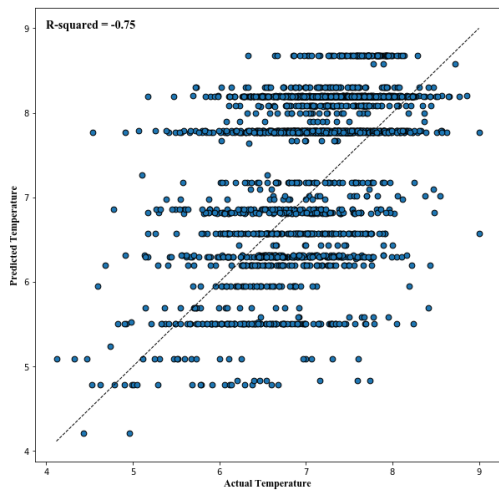| Dataset	Scenarios				
	1	2	3	4	5
Training	Montreal	Apeldoorn	Montreal	Apeldoorn	Combined
Testing	Montreal	Apeldoorn	Apeldoorn	Montreal	Combined

In scenario one, the dataset consisted of 5,578 data instances distributed in 4,183 feature vectors for training, and 1,395 feature vectors for testing. The model performs with a Mean Absolute Percentage Error (MAPE) of 0.07, and an R-squared (R^2) of 0.61 for the testing data sets. Figure 5(a) presents the overall dispersion of the predictions in terms of the predicted temperatures.
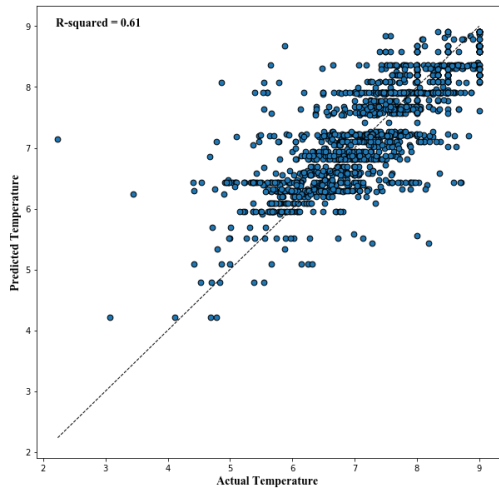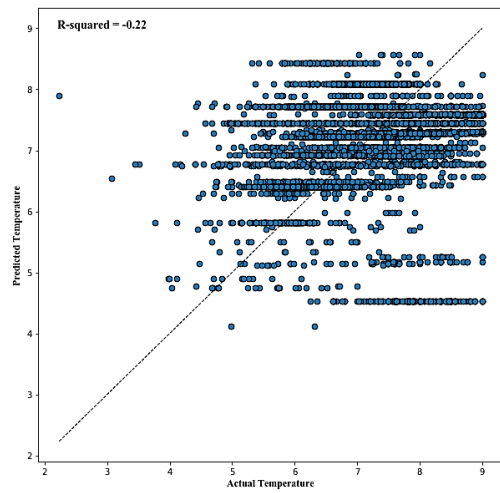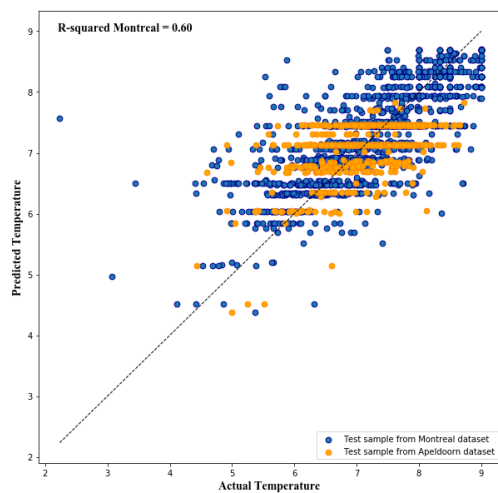
(a) Scenario 1
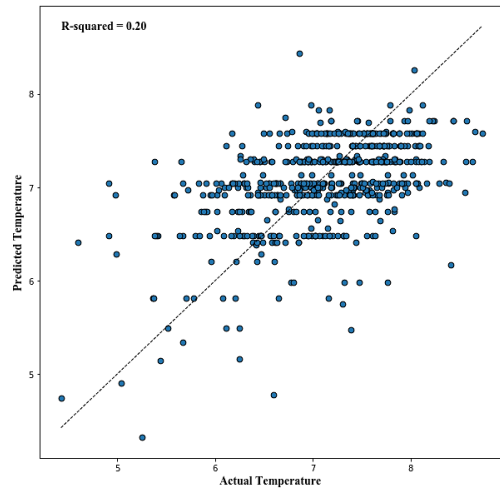
(b) Scenario 2

(c) Scenario 3

(d) Scenario 4

(d) Scenario 5

Figure 5. Model performances for different scenarios scenario 5

In scenario two, the training and testing datasets consist of 1806 and 602 instances, respectively. The model performs with a MAPE of 0.07 and a R^2 of 0.20 for the testing data samples. Figure 5(b) shows the accuracy in terms of the predicted temperatures. The models show a higher dispersion, not fitting very well the predicted values vs. the actual temperature values. On the other hand, for scenario three, the already trained model from scenario one (i.e., the Montreal model) was tested with the dataset from Apeldoorn. Figure 5(c) shows the performance of the model, with a MAPE of 0.11 and an R-squared of -0.75, for the testing data set. Similarly, in scenario four, the trained model from Apeldoorn, is tested with data from Montreal. Here the MAPE is 0.12 and the R-squared is -0.22. Figure 5(d) shows the performance of the model. In the last scenario, the two data sets (i.e., Apeldoorn and Montreal) are used together to train the model. The dataset consists of 7,986 instances. A total of 5,989 data instances were used to train the model (75% of the total dataset). To test the accuracy of the model, a sample of 1,395 data instances from Montreal and 602 from Apeldoorn were selected. The model performs with a MAPE of 0.07, and a R-squared of 0.60 for the testing data samples, showing better generalizability and improved performance compared to scenarios two through four. Figure 5(e) shows the dispersion of the model.

4   Discussion

From the results provided by the scenarios evaluated in this research, two main discussion points can be derived.

(1) As presented in Table 2, scenarios one and five performed the best, with a MAPE of 0.07 for both models, and an R-squared of 0.61 and 0.60, respectively. This highlights the importance of diversity in the datasets. In the particular case of scenario one, given the very nature of the city of Montreal as a metropolitan city with much wider variety in urban morphology and socio-economic characteristics, the model is based on a more diverse dataset. Therefore, it is justified that the model performs better than when compared to models trained with smaller datasets, as in the case of scenario two. The same can be observed in scenario five, where the model has data samples from both cities, and therefore learns from different contexts with a good performance.

However, when the models are faced with the challenge of predicting data from different urban environments and outside the context of the training data (i.e., scenarios three and four), both models perform poorly. In the case of the model from Montreal when tested on the data from Apeldoorn (i.e., scenario 3), while

the MAPE is comparatively good (as highlighted in Figure 6, most of the data are in the range of 0 to 14 % error range); however R^2 is very low. Negative R^2 in Scenarios 3 and 4 suggests that the models perform worse than taking the mean temperature class value. Nevertheless, R^2 is not always a good indicator of fitness, depending on the characteristics of the data [23, 24]. According to these studies, the performance of the model needs to be evaluated in the context of how well the predictions are made on unseen datasets, (a.k.a., the predictive capacity of the model) and not only in goodness of fit. Whether or not this is the case in this research, still needs to be investigated. The same can be said for scenario four, where the estimated error is 12 %.

Table 2. Results from scenarios studied in this research

Performance Metric	Scenarios				
	1	2	3	4	5
MAPE	0.07	0.07	0.11	0.12	0.07
R-squared	0.61	0.20	-0.75	-0.22	0.60

(2) A very recurrent problem during this research has been the inconsistency of what is understood by UHI in different urban contexts. For instance, in the case of Montreal, a tailored thermal classification of hot and cold islands was generated. This was done by using a statistical model with geospatial variables generated from a Landsat 8 satellite image. Variables such as: (1) normalized difference vegetation index (NDVI), (2) the impermeability index, (3) the average air temperature on the day and in the 72 hours prior to the acquisition of the LST images, and (4) the average wind speed on the day the images were acquired for the classification [25]. In the case of Apeldoorn, the assessment is based on PET-model, where it is implemented for a standardized person (male, 35 years old, 1.75 m tall and weighing 75 kg) for 10 typical Dutch street configurations. In this case, the vegetation index (NDVI) was also included, in addition to the land use, Sky View Factor (SVF), and daily air temperatures [21].

Although the two heat maps assist different governments and urban planners in designing their mitigation strategies, and ways to deal with UHI in their own contexts, the two heat maps are not comparable, and thus, create a different understanding of the UHI phenomenon, and corresponding mitigation strategies in the two countries.

The current trend in urban planning is gravitating towards the use of digital twins and data-driven methods. These technological advances provide cities with a unique opportunity to evaluate UHI in a systematic and consistent manner worldwide, at both local and global

scales, and therefore feeding knowledge and innovation from one context to another. Yet, the heterogeneous and inconsistent manner in which local governments and municipalities assess UHI hampers the application of data-driven methods. This is because the training dataset per se can have different meanings across urban environments.
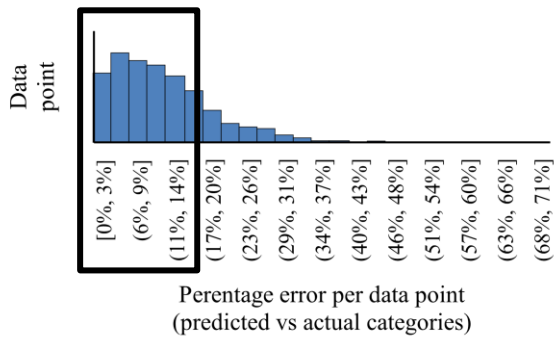

Figure 6. Percentage error in scenario 3, where the majority of the data points have 0- 15% error.

5   Conclusion and future work

This study evaluated the extent to which data-driven UHI assessment models can be generalized to different urban contexts by applying the previously data-driven UHI assessment framework proposed by the authors to the cities of Montreal, Canada, and Apeldoorn, the Netherlands, in five different scenarios.

The results of the five scenarios show the potential of data-driven models to predict UHI categories with a fair accuracy. However, it was also highlighted that while each model has good prediction capabilities within the scope of its specific training data set, these models do not demonstrate good generalizability on the testing data from a different context (i.e., city).

It should be mentioned that the temperature datasets where centered mainly in categories 5 to 8, with small samples from lower, and higher categories. This has created an imbalanced dataset. In the future, efforts are required to sample wider data points from all categories.

An interesting future research would be to investigate the causes of variability in the behavior of UHI of different urban contexts. The previous work of the authors have considered this for the single case of Montreal [14]; but it is hypothesized that the important features that govern the behavior of UHI heavily depends on the urban context. In the future, the authors would like to expand the scope of this investigation and study the most dominant causes of UHI variability between different cities.

Moreover, as cities continue to grow, so does the need for a standardized understanding of UHI at a pedestrian level. The presented research highlights the need to develop standardized methods for definition and measurement of UHI. The intrinsic differences in the way UHI is evaluated and dealt with can potentially create confusion about the phenomenon across the world, and constraint the applicability and generalizability of data-driven approaches.

References

[1]   L. Howard, "The Climate of London," 1833.

[2]   J. Tan *et al.*, "The urban heat island and its impact on heat waves and human health in Shanghai," *International journal of biometeorology,* vol. 54, no. 1, pp. 75-84, 2010.

[3]   P. Acharya, B. Boggess, and K. Zhang, "Assessing heat stress and health among construction workers in a changing climate: a review," *International journal of environmental research and public health,* vol. 15, no. 2, p. 247, 2018.

[4]   Climate Adapt. (2021, 01 June).   [Online]. Available: https://climate-adapt.eea.europa.eu/knowledge/tools/urban-adaptation/climatic-threats/heat-waves.

[5]   United Nations, "World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420)," New York: United Nations, 2019, [Online]. Available: https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf.

[6]   T. Sutjaritvorakul, A. Vierling, and K. Berns, "Data-Driven Worker Detection from Load-View Crane Camera," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, 2020, vol. 37: IAARC Publications, pp. 864-871.

[7]   A. K. Langroodi, F. Vahdatikhaki, and A. Doree, "Activity recognition of construction equipment using fractional random forest," *Automation in Construction,* vol. 122, p. 103465, 2021.

[8]   A. Nutkiewicz, Z. Yang, and R. K. Jain, "Data-driven Urban Energy Simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow," *Applied energy,* vol. 225, pp. 1176-1189, 2018.

[9]   S. Yoo, "Investigating important urban characteristics in the formation of urban heat islands: a machine learning approach," *Journal of Big Data,* vol. 5, no. 1, pp. 1-24, 2018.

[10] S. Vulova, F. Meier, D. Fenner, H. Nouri, and B. Kleinschmit, "Summer nights in Berlin, Germany: modeling air temperature spatially with remote sensing, crowdsourced weather data, and machine learning," *IEEE Journal of Selected Topics in*

Applied Earth Observations and Remote Sensing, vol. 13, pp. 5074-5087, 2020.

[11] Y. Sun, C. Gao, J. Li, R. Wang, and J. Liu, "Quantifying the effects of urban form on land surface temperature in subtropical high-density urban areas using machine learning," *Remote Sensing,* vol. 11, no. 8, p. 959, 2019.

[12] Y. Kwak, C. Park, and B. Deal, "Discerning the success of sustainable planning: A comparative analysis of urban heat island dynamics in Korean new towns," *Sustainable Cities and Society,* vol. 61, p. 102341, 2020.

[13] K. Gobakis, D. Kolokotsa, A. Synnefa, M. Saliari, K. Giannopoulou, and M. Santamouris, "Development of a model for urban heat island prediction using neural network techniques," *Sustainable Cities and Society,* vol. 1, no. 2, pp. 104-115, 2011.

[14] M. Pena Acosta, F. Vahdatikhaki, J. Santos, A. Hammad, and A. G. Dorée, "How to bring UHI to the urban planning table? A data-driven modeling approach," *Sustainable Cities and Society,* vol. 71, p. 102948, 2021/08/01/ 2021.

[15] M. Parsaee, M. M. Joybari, P. A. Mirzaei, and F. Haghighat, "Urban heat island, urban climate maps and urban development policies and action plans," *Environmental technology & innovation,* vol. 14, p. 100341, 2019.

[16] "Statistics Netherlands Land Use 2015," pdok, Ed., ed, 2015.

[17] M. Bondarenko, D. Kerr, A. Sorichetta, and A. Tatem, "Census/projection-disaggregated gridded population datasets for 189 countries in 2020 using Built-Settlement Growth Model (BSGM) outputs," 2020.

[18] "3D Basisvoorziening," pdok, Ed., ed, 2019.

[19] "Nationaal Wegenbestand (NWB)," NWB, Ed., ed, 2021.

[20] Gehl - Making Cities for People. Public Spaces and Public Life, Apeldoorn 2009. Available: https://issuu.com/gehlarchitects/docs/apeldoorn_10 0616

[21] S. Koopmans, B. Heusinkveld, and G. Steeneveld, "A standardized Physical Equivalent Temperature urban heat map at 1-m spatial resolution to facilitate climate stress tests in the Netherlands," *Building and Environment,* vol. 181, p. 106984, 2020.

[22] G. F. Jenks, "The data model concept in statistical mapping," *International yearbook of cartography,* vol. 7, pp. 186-190, 1967.

[23] C. Onyutha, "From R-squared to coefficient of model accuracy for assessing "goodness-of-fits"," *Geosci. Model Dev. Discuss.,* vol. 2020, pp. 1-25, 2020.

[24] S. Rose and T. G. McGuire, "Limitations of p-values and R-squared for stepwise regression building: a fairness demonstration in health policy risk adjustment," *The American Statistician,* vol. 73, no. sup1, pp. 152-156, 2019.

[25] E. Boulfroy, J. Khaldoune, F. Grenon, R. Fournier, and B. Talbot, "Conservation des îlots de fraîcheur urbains - Description de la méthode suivie pour identifier et localiser les îlots de fraîcheur et de chaleur (méthode en 9 niveaux)," CERFO et Université de Sherbrooke, 2013.