Utilizing World Urban Database and Access Portal Tools (WUDAPT) and machine learning to facilitate spatial estimation of heatwave patterns

Yuan Shi [a], Chao Ren [b], Ming Luo [c], Jason Ching [d], Xinwei Li [b], Muhammad Bilal [e], Xiaoyi Fang [f] *, Zhihua Ren [g]

[a] Institute of Future Cities (IOFC), The Chinese University of Hong Kong, Hong Kong SAR, China

[b] Faculty of Architecture, The University of Hong Kong, Hong Kong SAR, China

[c] School of Geography and Planning, Sun Yat-Sen University, Guangzhou, China

[d] Institute for the Environment, University of North Carolina, Chapel Hill, NC, USA

[e] School of Marine Sciences, Nanjing University of Information Science & Technology, Nanjing, China

[f] Chinese Academy of Meteorological Sciences, China

[g] National Meteorological Information Center, China

The corresponding author's* email addresses: fangxy@cma.gov.cn

Highlights

- Heatwave events and hot day numbers were investigated for a subtropical region;

- Random forest was adopted to facilitate spatial estimation of heatwave patterns;

- Utilizing WUDAPT as an input of machine learning-based urban climate studies;

- WUDAPT are proven to be useful predictor variables of heatwave estimation;

- Spatial patterns of heatwave were mapped at a high spatial resolution.

1 **Utilizing World Urban Database and Access Portal Tools**

2 **(WUDAPT) and machine learning to facilitate spatial estimation**

3 **of heatwave patterns**

4

5 **Abstract**

6 Climate change lead to more intense, higher frequent and prolonged heat extremes.

7 Understanding the spatial pattern of heatwave is vital for providing the corresponding

8 weather services, making climate change adaptation strategies and heat-health actions. In this

9 study, we present an approach to estimate the heatwave spatial patterns by utilizing the

10 WUDAPT Level 0 data and machine learning. The analysis is based on two years (2009 and

11 2016) of air temperature data from 86 meteorological monitoring stations in Guangdong

12 province of China, a subtropical region with frequent hot and sultry weather in summer. First,

13 heatwave conditions were quantified by calculating the number of hot days and frequency of

14 heatwave events in each year and used as the response variables. Then, random forest models

15 were built by using a geospatial dataset consisting of WUDAPT and urban canopy

16 parameters (UCP) as predictor variables. Based on the resultant models, spatial patterns of

17 heatwave were estimated and mapped at 100m spatial-resolution. The results show that this

18 approach is able to estimate heatwave spatial patterns using open data and inform urban

19 policy and decision-making. The study is also a new perspective and a feasible pathway of

20 utilizing WUDPAT Level 0 product to facilitate urban environment applications.

21

Highlights

- Heatwave events and hot day numbers were investigated for a subtropical region;

- Random forest was adopted to facilitate spatial estimation of heatwave patterns;

- Utilizing WUDAPT as an input of machine learning-based urban climate studies;

- WUDAPT are proven to be useful predictor variables of heatwave estimation;

- Spatial patterns of heatwave were mapped at a high spatial resolution.

Keywords

Heatwave; WUDAPT; Random forest; Machine learning; Spatial estimation

1. Introduction

Climate change has been identified as a major challenge to environmental sustainability, human health and well-being (IPCC, 2014; WMO and WHO, 2015). In the context of climate change and the trend of global warming, heatwave events during summertime have become one of the most severe meteorological disasters in cities and societies. Heatwave generally refers to the extreme events of a period of consecutive hot weather (Meehl and Tebaldi, 2004; Nairn and Fawcett, 2011). In the past two decades, heatwave events during summertime become more intense, more frequently-happened, and longer-lasting (Field, 2012; Meehl and Tebaldi, 2004; Stocker, 2014). Such extremely hot weather conditions have brought serious negative impacts on environmental health (Haines et al., 2006) and also considerable economic loss (Epstein and Mills, 2005).

1.1. The associations of negative health outcomes and vulnerabilities with heatwave

There have been many studies emphasizing the associations of negative heat-health outcomes (Campbell et al., 2018; Mayrhuber et al., 2018). It is generally known that the elderly and

44 children were more susceptible and vulnerable to heatwave events (Benmarhnia et al., 2015;

45 Oudin Åström et al., 2011). However, extremely hot weather causes a series of heat-related

46 health impact ranging from sleeping disorder, to heat morbidity and event to the death, not

47 only for the elderly, children, and vulnerable peoples who do not have strong resistance

48 (Bunyavanich et al., 2003; Kenny et al., 2010; Maughan, 2012; Xu et al., 2012) but also

49 increases health risks for youth and working populations (Xiang et al., 2013). Long-term

50 exposure to weather with high ambient temperature could even endanger human's life (Rey et

51 al., 2009) and cost extra economic loss (García-Herrera et al., 2010; WMO, 2011). A national

52 study in China indicates that a total of 4.5% (95% confidence intervals (CI): 1.4%−7.6%)

53 excess deaths were associated with heatwaves in south China (Ma et al., 2015).

54 **1.2. The effect of urbanization on heatwave and its spatial heterogeneity**

55 Globally, the process of urbanization is continuing (UN, 2018; UN, 2019). The influence of

56 urbanization on the spatiotemporal distribution of hot weather and the duration and intensity

57 of heatwave is noticeable, especially for those countries that are experiencing rapid

58 urbanization process (Guo et al., 2020; Oleson et al., 2015). The extremely hot weather

59 condition is exacerbated by urbanization (Chapman et al., 2017; Luo and Lau, 2016; Sun et

60 al., 2016). Generally speaking, urbanization increases the intensity of the urban heat island

61 (UHI) effect, which further increases the intensity, frequency, and duration of the heatwave

62 (Li and Bou-Zeid, 2013; Oke, 1973; Oke, 1997; Tan et al., 2010). In detail, the spatial

63 variability in land-use patterns and inhomogeneous land surface thermal and aerodynamic

64 properties lead to spatial heterogeneity in the near-ground wind field (Comrie, 2000),

65 radiation and energy balance (Arnfield, 2003), and anthropogenic heat (Taha, 1997). All

66 these spatial heterogeneities make the extremely hot weather condition varies among

67 locations (Hart and Sailor, 2009). Such spatial heterogeneity makes city dwellers who live in

68 urbanized areas, especially those who live in compact built-up areas in large cities are more

69 vulnerable to heatwave (Uejio et al., 2011a; WMO and WHO, 2015).

70 **1.3. The necessity of incorporating fine-scale spatial heterogeneity into the**

71 **estimation of heatwave condition**

72 Most of the heat-health related studies are based on time-series analysis; therefore, more

73 focuses on the temporal characteristics (frequency, duration) of heatwave events (Kovats and

74 Hajat, 2008). These studies usually take a city or an area of interest as a whole to associate

75 health burdens with temperature-related variables (the spatial extent ranges from several

76 kilometers to dozens of kilometers). Therefore, the effect of urbanization and the spatial

77 variability of heatwave weather have become missing elements in most of the above studies

78 and have not been comprehensively investigated (Kaiser et al., 2007; Kyselý, 2002; Le Tertre

79 et al., 2006). Many studies have demonstrated that the vulnerability of citizens to heatwave is

80 associated with demographic variables and socioeconomic factors (Bao et al., 2015; Chan et

81 al., 2012; Gronlund et al., 2015; Uejio et al., 2011b). Most of these studies overlay heat

82 exposure map, surface/air temperature map on the spatial information on vulnerable

83 population, in which spatial scale mismatch issues between the heatwave map and spatial

84 information on the corresponding vulnerability have been a concern (i.e., the Modifiable

85 Areal Unit Problem) (Fraser et al., 2018; Ho et al., 2015; Wong, 2004). Commonly-used

86 spatial units of the city administrative boundaries or urban planning zoning correspond to a

87 spatial resolution approximately a couple of kilometers. In such cases, the assessment of

88 urban vulnerability to heatwave events and the corresponding prevention measures would

89 have higher spatial uncertainties. The uncertainties could be even more considerable in those

90 regions with a complex geographic context.

91 The importance of locating the groups of people with a high-vulnerability to the impacts of

92 heatwave has been emphasized (Johnson et al., 2012), which means that it is vital to take

93 spatial elements into consideration. In that case, acquiring a detailed fine-scale spatial

94 understanding of the heatwave is essential to heat risk prevention and public health actions

95 (Buscail et al., 2012). In recent years, relevant studies have been conducted for the spatial

96 mapping of heat-related risks in many large or megacities worldwide (Dugord et al., 2014;

97 El-Zein and Tonmoy, 2015; Klein Rosenthal et al., 2014; Lemonsu et al., 2015; Wolf and

98 McGregor, 2013). Significant spatial variabilities of heat-related health impact were found in

99 all the above cases. Undoubtedly, reliable fine-scale spatial information on heatwave is a

100 fundamental part of all strategic actions in relation to the reduction of heat-related risk and

101 vulnerability. Moreover, it is also a vital part of the information for urban planning for a more

102 resilient built environment under heatwaves (Maragno et al., 2020).

103 **1.4. Study objective**

104 Understanding the spatial pattern of heatwave is important for public health management and

105 urban development. Despite the necessity of incorporating spatial heterogeneity into the

106 estimation of heatwave has been recognized, yet few studies focus on direct mapping of the

107 spatial distribution of heatwave conditions at a spatial resolution that finer than the

108 commonly-used spatial units of the city administrative boundaries or urban planning zoning.

109 To address the above research gap, in this study, we present an approach for the investigation

110 of the spatial heterogeneity of the heatwave condition via machine learning and geospatial

111 mapping techniques. The high-resolution spatial maps of heatwave generated by this study

112 will enhance the robustness of spatial assessment and mapping of heat-health risk. Besides

113 the resultant spatial mapping of the heatwave condition, the important influential variables of

114 heatwave will be identified by the study, which will also provide valuable clues for how the

115   cities should be properly planned to enhance the resilience to heatwave events and heat-

116   related disasters.

117   **2. Materials and methods**

118   In this study, spatial buffering analysis and machine learning technique - random forest were

119   adopted. Guangdong province in China, a subtropical region with highly diversified

120   landscape and spatially heterogeneous land surface coverage, was selected as the study area,

121   as its geographical complexity makes it an ideal testbed for the research. Land cover types

122   were analyzed by utilizing the Level 0 data product of World Urban Database and Access

123   Portal Tools (WUDAPT) (Bechtel et al., 2019; Bechtel et al., 2016). Land surface

124   morphology was also quantified by introducing urban canopy parameters (UCP). First, a

125   buffering method was adopted to analyze WUDAPT Level 0 data and UCP geospatial data

126   and generate predictor variables. Then, the influential predictors of the heatwave were

127   identified via random forest variable selection. Finally, the spatial pattern of heatwave was

128   estimated by performing random forest regression modelling. Figure 1 shows the workflow

129   of the present study. The year 2016 and 2009 that we selected are two of the representative

130   warmest years on the local records. Local records indicate frequent occurrence of

131   meteorological disasters related to heatwaves in these two years. In this study, the workflow

132   was performed twice to generate the heatwave maps for both of the years 2009 and 2016 in

133   which the WUDAPT land type data were also generated. Thus, it is feasible to explore the

134   temporal change of the effect of urbanization in this period of eight years. All methodological

135   details are introduced in the following paragraphs of this section.

136
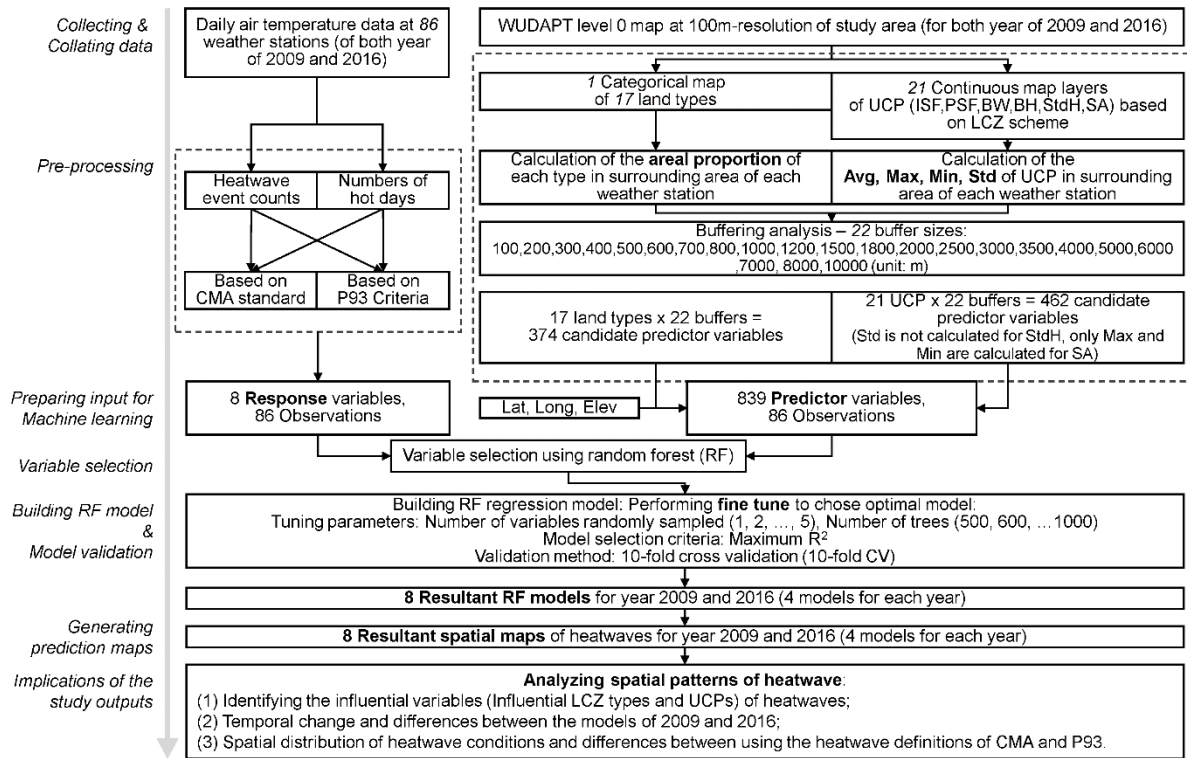
Figure 1. *The workflow of the present study.*
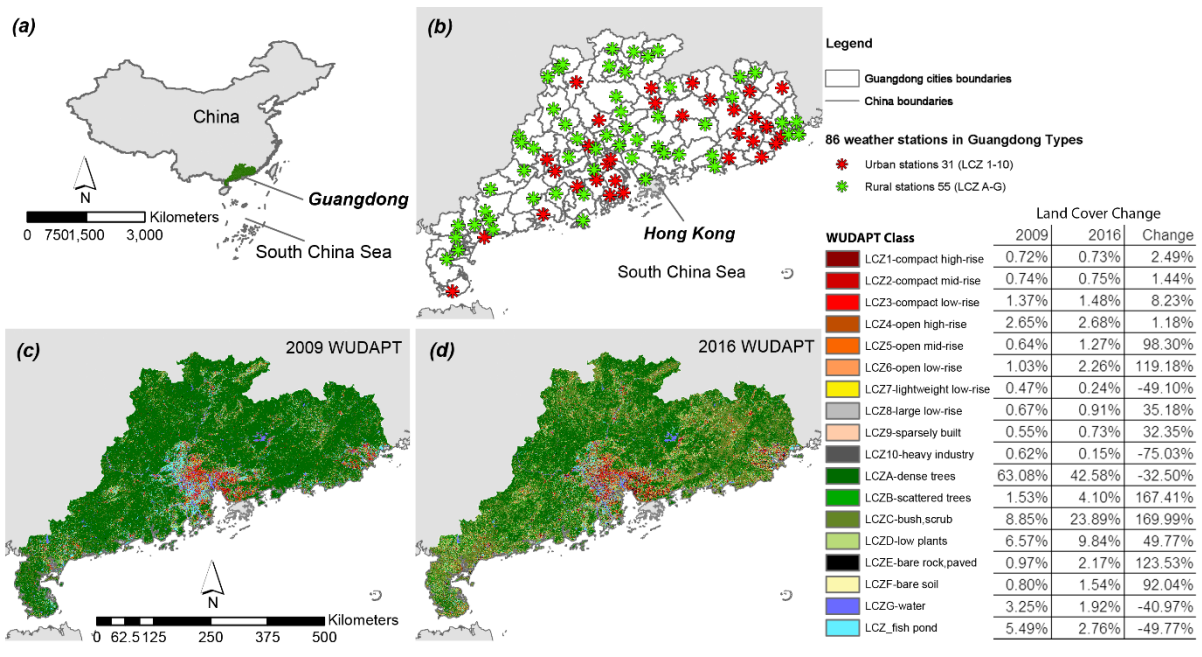
138   **2.1. Response variables**

139   **2.1.1. Heatwave definition**

140   There is no standardized definition for heatwave worldwide. Currently, China adopts the

141   standard of heatwave developed by the China Meteorological Administration (CMA). The

142   CMA defines a day as a high-temperature hot day when it has a maximum daily temperature

143   ≥ 35 °C, and a heat weather event consists of a consecutive three hot days or more than three

144   hot days is defined as a heatwave event. However, using a single and absolute definition for

145   heatwave investigations for China could possibly introduce bias, as China's vast territorial

146   area spans a wide range of latitudes and contains many different climatic zones. It has been

147   found that the peoples' group-specific mortality risks in a city in humid subtropical climate

148   zone (Köppen Cfa) in China using the standardized heatwave definition of CMA was

149   significantly underestimated than that using the heatwave definition of daily mean

150   temperature $\geq$ 99.0th percentile (P99) with a duration $\geq$ 3 days (Zhang et al., 2017). In that

151   case, a national-scale study on the adjustment of heatwave definition has emphasized the

152   importance of area-specific definitions of heatwave in heat-health risk assessments and

153   developed a series of area-specific heatwave definitions for different regions in China (Lin et

154   al., 2019). In the study of China, the regional heatwave of Northeast, North, Northwest, East,

155   Central and Southwest China were defined separately as being two or more consecutive days

156   with a daily mean temperature higher than or equal to the P64, P71, P85, P67, P75 and P77 of

157   the warm season (May to October) temperature. The heatwave in South China (for example,

158   Guangdong province) was defined as five or more consecutive days with a daily mean

159   temperature higher than or equal to the P93 of the warm season temperature. In the present

160   study, both the CMA heatwave definition and the P93 definition were adopted to define

161   heatwave. As a result, two groups of models will be generated based on two different criteria.

162   **2.1.2.   Heatwave event counts and the numbers of hot days**

163   There is a total of 86 national meteorological monitoring stations currently located in the

164   study area – Guangdong province (Figure 2). All the stations are operated by CMA and

165   conform to the WMO guide (WMO, 2008). Hourly air temperature data of the year 2009 and

166   2016 were collected and used for the analysis. Using CMA and P93 definition, the numbers

167   of hot days and heatwave event counts of the entire year were calculated for both of the two

168   years, respectively. Comprehensively considering the event counts and hot day numbers leads

169   to a holistic understanding of the heatwave characteristics about duration and frequency. As a

170   result, a total of eight parameters were calculated to represent the heatwave conditions and

171   used as the model response variables (2 definitions $\times$ 2 parameters $\times$ 2 years). Consequently,

172   eight models and corresponding prediction maps will be generated.

WUDAPT Class	Land Cover Change		
	2009	2016	Change
LCZ1-compact high-rise	0.72%	0.73%	2.49%
LCZ2-compact mid-rise	0.74%	0.75%	1.44%
LCZ3-compact low-rise	1.37%	1.48%	8.23%
LCZ4-open high-rise	2.65%	2.68%	1.18%
LCZ5-open mid-rise	0.64%	1.27%	98.30%
LCZ6-open low-rise	1.03%	2.26%	119.18%
LCZ7-lightweight low-rise	0.47%	0.24%	-49.10%
LCZ8-large low-rise	0.67%	0.91%	35.18%
LCZ9-sparsely built	0.55%	0.73%	32.35%
LCZ10-heavy industry	0.62%	0.15%	-75.03%
LCZA-dense trees	63.08%	42.58%	-32.50%
LCZB-scattered trees	1.53%	4.10%	167.41%
LCZC-bush,scrub	8.85%	23.89%	169.99%
LCZD-low plants	6.57%	9.84%	49.77%
LCZE-bare rock,paved	0.97%	2.17%	123.53%
LCZF-bare soil	0.80%	1.54%	92.04%
LCZG-water	3.25%	1.92%	-40.97%
LCZ_fish pond	5.49%	2.76%	-49.77%

Figure 2. *The location of the study area - Guangdong province in China (a). The location of 86 national meteorological monitoring stations in Guangdong province (b). The WUDAPT map of the year 2009 (c) and 2016 (d). A table (at the right side of the legend) to illustrate the LCZ difference between 2009 and 2016 and the rate of change.*

2.2. Predictor variables

The physical basis behind the spatial variability in air temperature has been comprehensively understood from the viewpoint of urban climatology (Landsberg, 1981; Oke, 1982; Oke, 1987; Oke, 1988). In the present study, to maximize the reproducibility and worldwide applicability of the workflow, all data used for generating predictor variables are open data based on worldwide unifiable standards. Two major parts of data were selected and used as the input for generating predictor variable datasets, which are: (a) Land surface cover – WUDAPT, (b) Urban canopy parameters (UCP), as they have been proved to be ideal proxies of land surface form as well as good indicators of the spatial variability in near-ground ambient air temperature (Bechtel et al., 2015; Salamanca et al., 2011).

2.2.1. Land surface cover – WUDAPT Level 0 map

WUDAPT Level 0 data has been popularly used for the investigation of spatial variability in air temperature (Leconte et al., 2015; Shi et al., 2018b). It is a well-established urban data portal which aims to provide a globally standardized and detailed urban morphological database of Local Climate Zone (LCZ) (Mills et al., 2015). LCZ is a standardized urban morphology scheme which provides a 17-LCZ type land surface classification for urban climate research (Stewart and Oke, 2012). Different LCZ types represent various combinations of surface structure (sky view factor, aspect ratio, surface roughness), surface cover (ground coverage ratio of buildings, vegetation, and impervious/paved surface), surface thermal properties, surface albedo, and human activity (building functions, anthropogenic heat). Moreover, different from other existing land use/land cover (LU/LC) classification products (e.g., USGS Global Land Cover Characterization (GLCC), Climate Change Initiative (CCI) Land Cover, and GlobeLand30), the WUDAPT introduces 3-level product to meet different needs in urban climate study and as well as provides a detailed LU/LC classification, especially for the built-enviroment. The above features make it a better proxy to depict the aerodynamic and thermal properties of the land surface. There are mainly two types of methods of generating LCZ map - GIS-based method and satellite image-based method (Gál et al., 2015; Wang et al., 2018). The GIS-based method is more city-specific as it uses local datasets. The robustness of GIS-based results depends on the quality of local urban datasets, thus varying from place to place. Oppositely, the satellite image-based method, as the most popular one of the WUDAPT Level 0 product methods, was designed to be universal to be part of a global dataset of urban form (Ching et al., 2018). Specifically, LCZ map at a high spatial-resolution of 100 meters (Level 0 data) can be generated by using open-source satellite images (Bechtel et al., 2015). As the input data, the highly standardized database of WUDAPT also enables cross-comparison between urban climate related studies

213   in different regions of the world. The above advantages make WUDAPT a superior choice of

214   input data to facilitate urban climate and environmental modelling (Ching et al., 2014). The

215   WUDAPT Level 0 map of Guangdong province (Figure 2) was generated for both of the

216   years 2009 and 2016 in previously published peer-reviewed research. The information on

217   accuracy assessment can be found in two papers. WUDAPT Map of 2009 is described in the

218   study by Wang et al. (2019b). WUDAPT Map of 2016 is described in the study Ren et al.

219   (2019).

220   ### 2.2.2. Urban canopy parameters (UCP)

221   There have been studies focus on the relationship between parameters of urban surface

222   parameterization such as the urban canopy parameters (UCP) and the ambient air temperature

223   or urban heat island effects (Chen et al., 2011; Garuma, 2018; Salamanca et al., 2011; Sharma

224   et al., 2017). In the present study, besides WUDAPT, six commonly used parameters were

225   also used for generating continuous data layers of predictor variables, which are impervious

226   surface fraction (ISF, also known as urban fraction), pervious surface fraction (PSF), building

227   width (BW), building height (BH), standard deviation of building heights (StdH), surface

228   albedo (SA). All building-related parameters were estimated from 30m resolution DSM and

229   DEM datasets by using the method developed by Ren et al. (2020). Specifically, the building

230   footprints are extracted by utilizing the Google Maps Static API. The height information is

231   estimated from the Advanced Land Observing Satellite (ALOS) World 3D Digital Surface

232   Model dataset. The 3D building morphology can be achieved by combing the above two parts

233   of information and be used to map UCP at a spatial resolution of 100m. Instead of

234   exhaustively processing data for the large spatial extent (approximately 179,800 km^2) of the

235   entire study area, an inexhaustive sampling strategy was used for generating the spatial maps

236   of the six parameters. Specifically, for each LCZ type in the study area, we randomly select

237   30 - 100 typical LCZ sample sites that locate separately from each other. The average (Avg),

238 maximum (Max), minimum (Min), and standard deviation (Std) values of building-related

239 parameters were calculated for all samples of each LCZ type and then assigned to each $100 \times$

240 100 m pixel based on their LCZ type. Similarly, minimum and maximum SA values were

241 assigned to each pixel based on the representative values of the LCZ scheme (Stewart and

242 Oke, 2012). ISF and PSF were based on the High-resolution Multi-temporal Mapping of

243 Global Urban Land product (Liu et al., 2018). The 2010 and 2015 data layers were used in

244 this study, as the product only provides data layer at a 5-year interval.

245 ### 2.2.3. Buffering Analysis

246 The measured air temperature depends on not only the physical environment at the location

247 of meteorological stations but also its surroundings at a longer spatial range (Kljun et al.,

248 2004; Konarska et al., 2016). Therefore, instead of training the model by directly using the

249 predictor data extract at those pixels in which the meteorological stations located, we perform

250 buffering analysis to generate predictor variable sets. Similar buffering analysis has been

251 adopted and found to be a useful way of investigating fine-scale spatial variability of air

252 temperature in several studies (Brandsma and Wolters, 2012; Johnson et al., 2020; Schatz and

253 Kucharik, 2014; Shi et al., 2018a). In this present study, a total of 22 buffer radius range from

254 100 meters to 10 kilometers was used for buffering analysis (For example, a buffer zone with

255 a radius of 500 m contains a total of 80 pixels on a data layer of 100 m resolution). For the

256 categorical data of each of the 17 LCZ types, the areal proportion within the range of each

257 buffer radius was calculated. For all UCP parameters, the average value within the range of

258 each buffer radius was calculated based on the continuous data layers of UCP (mentioned in

259 section 2.2.2). All predictor variables were calculated for both the year of 2009 and 2016.

260 The above process, along with the latitude, longitude, and elevation of meteorological

261 stations, resulted in a total of 839 candidate predictors, which correspond to 839 spatial data

262 layers for each year. To be consistent with the WUDAPT maps, all spatial data layers were

263   generated using the resolution of 100×100 m; thus all have a data amount of 46.1

264   Megapixel. The prediction map will be using the same spatial resolution which is a relatively

265   fine resolution with regards to the spatial extent of the study area and allows to spatially

266   continuous data layer of heatwave patterns. All geospatial data processing was completed in

267   QGIS Desktop software (v3.10.5 LTR).

268   **2.3. Variable selection and regression modelling using random forests**

269   It has been found that non-linear modelling techniques are necessary for the estimation of

270   spatial variability in air temperature, as they usually result in a better prediction performance

271   than linear modelling approaches (Brandsma and Wolters, 2012; Voelkel and Shandas, 2017).

272   Ensemble approaches such as random forest allow building complex non-linear models while

273   still provide reasonable interpretability by ranking variable importance. In this study, the

274   random forest algorithm (Breiman, 2001) was used for both the variable selection and the

275   regression modelling.

276   **2.3.1. Predictor variable selection**

277   As the predictor variable dataset contains a massive amount of candidate predictor variables

278   that have to be examined, a random forest-based strategy of variable selection developed by

279   Genuer et al. (2010) was adopted by the present study for variable selection. Simply

280   speaking, the method ranks all the candidate predictor variables via the random forest

281   permutation-based score of importance, and during the process a forward stepwise strategy is

282   employed for adding predictor variables (Genuer et al., 2015). This variable selection method

283   identifies two subsets of important predictor variables: a larger subset of variables aims to all

284   possible interpretation but with redundancy in explaining the variability in the response

285   variable and a smaller subset of predictors aims to a more robust prediction without

286   redundancy. In this study, we use the later subset, as the study aims to prediction mapping.

287   The above variable selection process was performed in R (v3.6.3) using the VSURF package

288   (v1.1.0) (Genuer et al., 2015) and finished using data-driven default values.

289   ### 2.3.2.   Random forest regression modelling and model fine-tuning

290   Random forest is a supervised machine learning algorithm that uses the ensemble learning

291   method and uses the Out-of-bag (OOB) error to measures the prediction error (Breiman,

292   1996; Breiman, 2001; Liaw and Wiener, 2002). The prediction performance of the random

293   forest regression model is sensitive to parameter tuning (Probst et al., 2019). The number of

294   trees (ntree) and the number of variables considered for splitting at each node (mtry) are two

295   commonly considered tuning parameters. In this study, a fine-tuning process is employed to

296   let the random forest algorithm to automatically choose the optimal prediction model.

297   Specifically, an extended tune grid was set to automatically repeat the experiment with all

298   possible combinations between mtry ranging from 1 to 5 and ntree ranging from 500 to 1000

299   (using an increment of 100). In order to avoid the overfitting issue and evaluate the regression

300   model, the resampling process was done by using repeated 10-fold cross-validation (Burman,

301   1989). All response data were randomly divided into ten subsets, with nine subsets used as

302   the training dataset and the other one subset used as validation datasets. This process was

303   repeated ten times until all data have been used as validation data once. The coefficient of

304   determination (R^2) is selected as the metric to determine the optimal model. The above

305   random forest regression and fine-tuning process were performed in R using the caret

306   package (v6.0-86) (Kuhn, 2008).

3. Results

3.1. Predictor variable selection results

Following the method mentioned in section 2.3.1, for each of the eight response variables, a subset of important predictor variables was identified. The identified variables for each response are summarized and shown in Table 1.

Table 1. Summary of important predictor variables identified by the variable selection process and the formula of eight prediction models. For the nomenclature of predictor variables: LCZ8_1800 means the areal fraction of LCZ8 within the buffer radius of 1800 meters; SA_Max_2000 means the averaged value of the maximum surface albedo in all pixels within the buffer radius of 2000 meters, so on and so forth.

Year	Heatwave definition	Model formula
2009	CMA	*Number of hot days ~ Lat + Elev + LCZE_8000 + LCZE_7000 + Lon + LCZG_10000 + LCZD_6000+ LCZ8_1800 + LCZ8_2500*
		HW events count ~ Elev + Lat + LCZE_8000 + LCZE_7000 + Lon + LCZ8_1800 + LCZA_8000
	P93	*Number of hot days ~ LCZ8_8000 + LCZ8_7000 + LCZ6_10000 + LCZ8_5000 + LCZA_8000*
		HW events count ~ LCZ8_10000 + LCZ8_8000 + LCZ8_7000 + LCZG_200 + LCZG_10000 + LCZ6_10000 + SA_Min_10000
2016	CMA	*Number of hot days ~ Lat + LCZE_8000+ LCZ2_600 + Elev + LCZ3_300 + LCZF_8000 + ISF_Std_200*
		HW events count ~ Lat + Lon + LCZE_8000
	P93	*Number of hot days ~ LCZ1_5000 + LCZC_1500 + SA_Min_2000 + LCZ2_2000 + SA_Max_1500*
		HW events count ~ LCZ1_5000 + LCZC_1500 + LCZ1_400 + SA_Max_2000 + LCZ1_6000 + SA_Min_2000 + LCZG_6000 + LCZG_2500 + LCZG_3000 + LCZ2_2000 + LCZ5_200

3.2. Resultant models and evaluation

Following the method mentioned in section 2.3.2, the R^2, mean absolute error (MAE) and root mean square error (RMSE) were calculated for the model evaluation. Figure 3 shows the results and the comparison between the predicted and monitored values of the number of hot days and heatwave event counts based on CMA and P93 heatwave definitions, in the year of

323   2009 and 2016, respectively. The eight resultant models explain 47.8% to 70.3% of the

324   spatial variability of heatwave conditions. Most of the resultant models can explain

325   approximately 60% of the spatial variability. Figure 4 shows the variable importance

326   measured and ordered based on Mean Decrease Accuracy (%IncMSE) and Mean Decrease
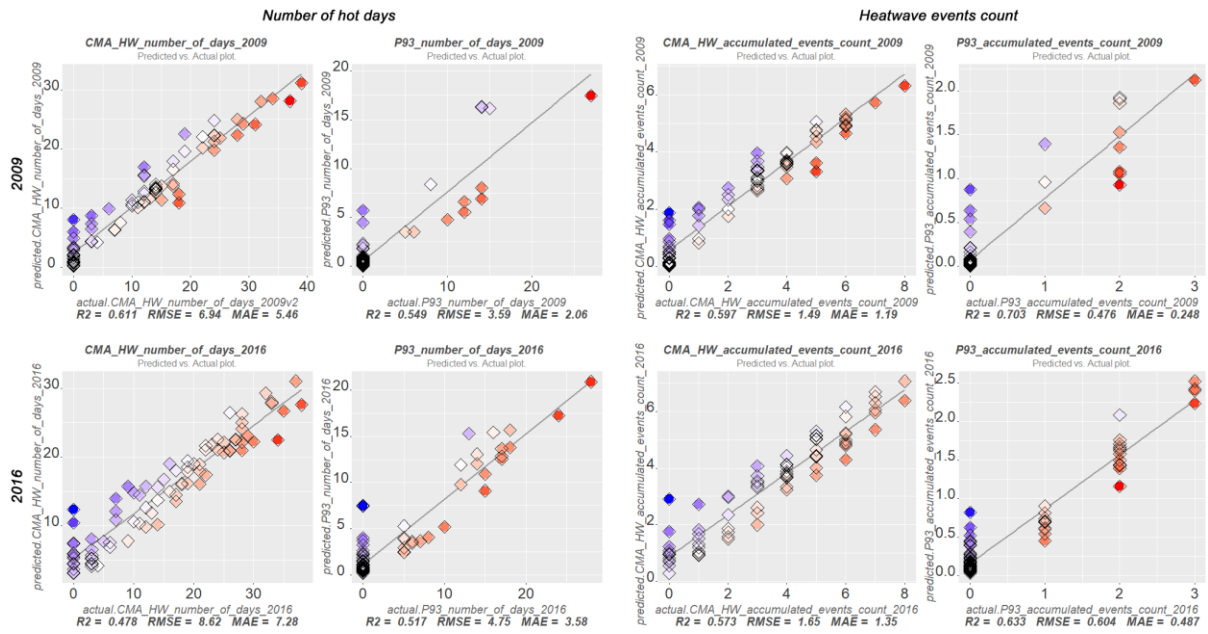
327   Gini (IncNodePurity).

328

Figure 3. *The actual-by-predicted data plot of the number of hot days and heatwave event counts based on CMA and P93 heatwave definitions, in the year of 2009 and 2016, respectively. The R^2, MAE, RMSE of corresponding models are shown under each plot.*
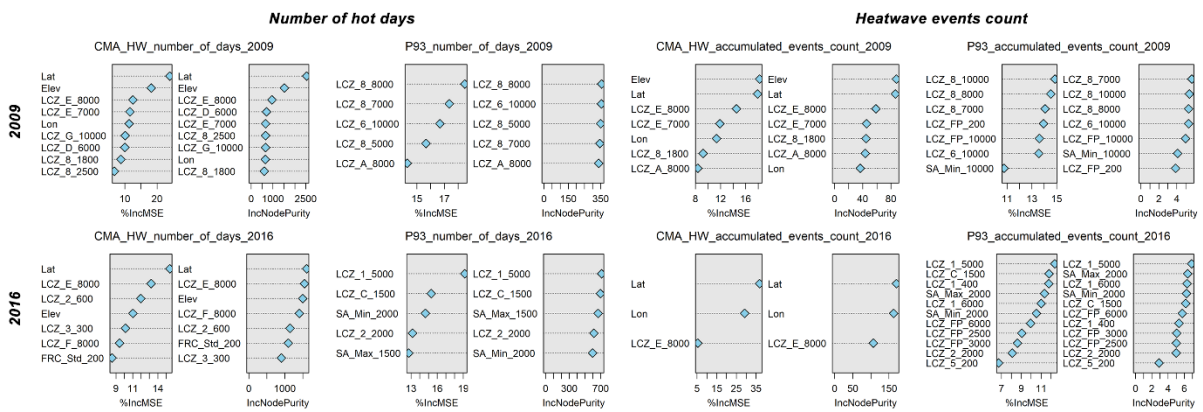

332

Figure 4. *The variable importance plot (%IncMSE and IncNodePurity sorted decreasingly from top to bottom) of all resultant models.*

335   **3.3. Spatial mapping of heatwave conditions**

336   Using the model formulas shown in Table 1 and the predictor data layers described in section

337   2.2.3, we generate the spatial prediction maps of the number of hot days and heatwave event

338   counts. As a result, eight maps were generated using the resolution of 100×100 m for the

339   entire spatial extent of the study area (Figure 5 and Figure 6).
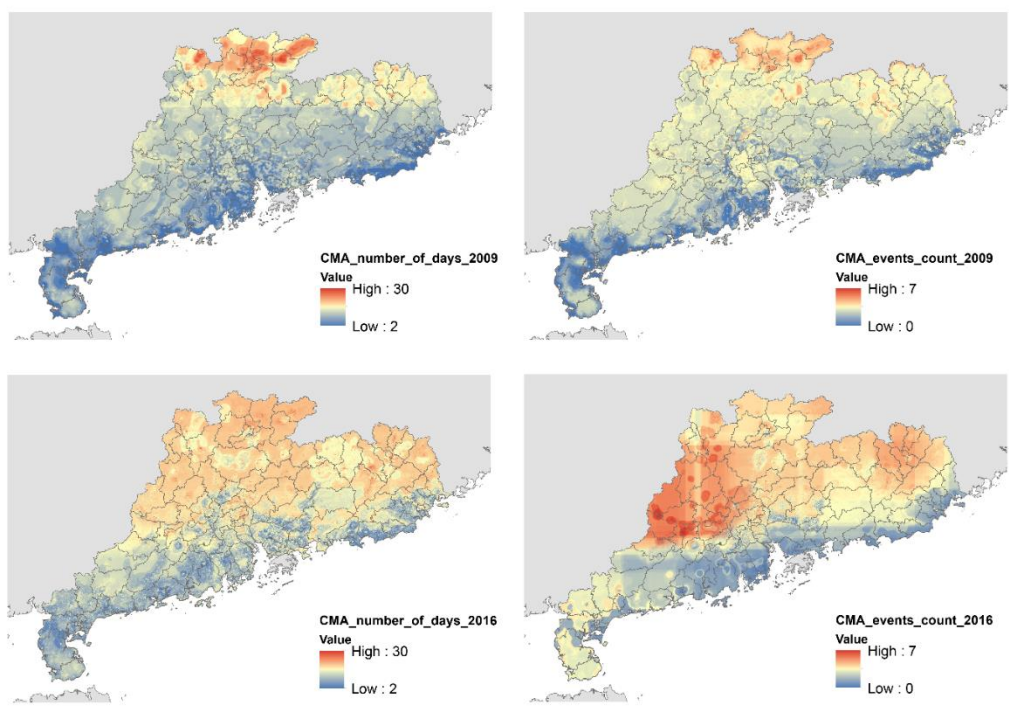
340

341   *Figure 5. Spatial prediction maps of the number of hot days and heatwave event counts*

342   *based on CMA heatwave definitions, in the year of 2009 and 2016, respectively. For the*

343   *number of hot days, only those days belongs to CMA heatwave events are counted. For*

344   *example, if a day has a maximum temperature ≥ 35 ℃, but not belongs to a ≥ 3-day*

345   *consecutive series, then it is not counted.*

346
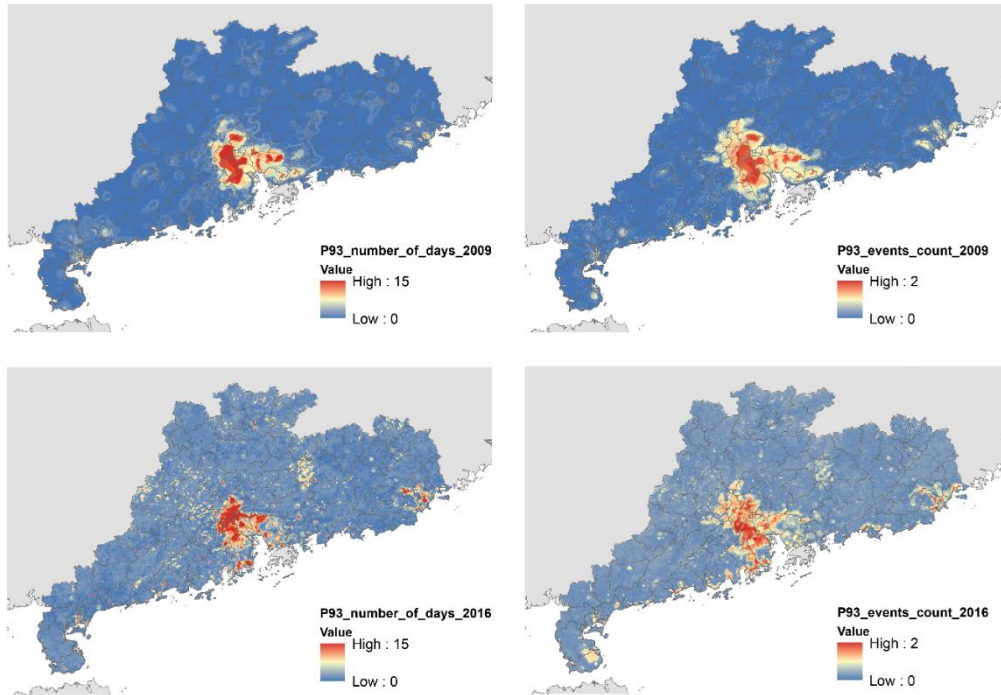
Figure 6. *Spatial prediction maps of the number of hot days and heatwave event counts*

based on P93 heatwave definitions, in the year of 2009 and 2016, respectively.

4. Discussion

4.1. Influential LCZ types and UCPs

In the eight resultant models, some LCZ types have been found that are commonly included,

which are LCZ 8 (large low-rise buildings), LCZ E (bare rock, paved surface), LCZ G (water

body). Besides the above LCZ types, LCZ 6 (open low-rise buildings) and LCZ A (dense

trees) were also included by some of the resultant models. Two UCPs were found to be

influential factors, i.e., impervious surface fraction, surface albedo. Impervious ground

surfaces in LCZ E, especially those artificial surfaces such as the pavements of concrete,

asphalt store more heat compared to natural pervious land surfaces (Arnold and Gibbons,

1996). Surface albedo has been proven to be a crucial factor in urban air temperature (Susca

et al., 2011; Taha et al., 1988). In the urban context of China, LCZ 8 mostly contains those

buildings with a very large area of flat dark-colored rooftop of low surface albedo, or the

361 color-coated steel sheet roofing. The thermal properties of roofing materials determine the

362 surface energy balance, consequently alters the atmospheric heating, thus affect ambient air

363 temperature (Coutts et al., 2013). LCZ A and LCZ G represent dense tree-dominated green

364 space and blue space. The blue-green space has strong and complex relationships with the

365 urban canopy and boundary-layer temperatures and has a significant impact on it (Garuma,

366 2018; Gunawardena et al., 2017; Morris et al., 2016). Therefore, it is reasonable that the

367 above candidate predictor variables are identified by the data-driven variable selection

368 process and included in the resultant models.

369 **4.2. Temporal change of the models between 2009 and 2016**

370 Comparing the 2009 and 2016 models, we found that LCZ1 (compact high-rise buildings),

371 LCZ2 (compact mid-rise buildings), and LCZ3 (compact low-rise buildings) which were not

372 included in the 2009 resultant models are included in the 2016 resultant models. These LCZ

373 types represent highly-urbanized compact built-up areas. During 2009 – 2016, Guangdong

374 province continues a rapid development and urbanization process, as planned in the "PRD

375 Region Reform and Development Planning Guidelines (2008–2020)" which was released by

376 the local authorities (Shen and Kee, 2016; Yu-shek, 2018). Affected by the regional planning

377 and development policies, the land area for urban development and construction is gradually

378 expanding, consequently causes an increase in local temperature. In cities, the amount of

379 high-rise buildings is increasing rapidly, and the building density is also increased. Such

380 changes in the urban morphology are conducive to the intensification of urban heat islands.

381 The inclusion of predictors about LCZ 1, LCZ2, and LCZ3 in the 2016 models implies that

382 the effect of urbanization on heatwave conditions becomes stronger during the past several

383 years. Such impacts of urbanization on local climate is also found in regional weather

384 simulation for Guangdong (Tse et al., 2018).

4.3. Spatial distribution of heatwave conditions and differences between using the heatwave definitions of CMA and P93

As shown in Figure 5 and Figure 6, the resultant maps based on CMA and P93 heatwave definitions have noticeable differences in the spatial pattern. In the spatial prediction maps based on the CMA definition, it is observed that the total number of hot days and heatwave events counted in coastal areas of Guangdong province are less than those in the inland areas. This finding breaks through the common mindset that the higher the average temperature, the more heatwave events. Using the absolute definition like CMA definition, the spatial distribution of hot days/heatwave events does not have to be consistent with the climatological mean temperature distribution. In South China, it has been found that a consecutive series of several hot days have higher health impacts than a single hot day with a extremely high temperature, as people cannot be relieved and physically recover from the hot weather (Wang et al., 2019a). Therefore, the heatwave investigation consider not only the air temperature but also the duration of the events. In this study, for the number of hot days, only those days belongs to CMA heatwave events are counted. For example, if a day has a maximum temperature ≥ 35 ˚C, but not belongs to a ≥ 3-day consecutive series, then it is not

counted. This has brought a counter-intuitive result that there are more heat waves in the northern part of the study area. It is also found that the number of hot days and heatwave events have a significant correlation with latitude and longitude (particularly in the map of heatwave events count in 2016). It is commonly known that the air temperature in a specific geographic location highly depends on the latitude and whether the location is near to the coastal area. The CMA definition uses a fixed threshold of air temperature (a maximum daily temperature ≥ 35 °C) to define hot days. Therefore, it is reasonable that the spatial distribution of heatwave conditions based on CMA definition is strongly correlated with geolocation predictors (latitude and longitude). Using a fixed threshold of air temperature to

410 investigate the spatial pattern of heatwave conditions in a relatively large spatial extent could

411 introduce bias. Specifically, it might overestimate the heatwave condition and relevant

412 environmental and health risks in low latitudes and underestimate the situation in high

413 latitudes, as the surface air temperature is already a function of latitude, and those people live

414 in different latitudes has varying tolerance and adaptions to heat (McCarthy et al., 2001).

415 Unlike CMA based results, the effect of urbanization can be clearly observed in the resultant

416 prediction maps produced using P93 definition. The hotspot shows in Figure 6 is the location

417 of Guangzhou which is the capital city and also the most urbanized and the most populated

418 city of Guangdong province. It has been found that city areas do have more very hot days

419 than rural areas due to the high-density building clusters trap heat within the city during

420 nighttime (Shi et al., 2019). Therefore, it is reasonable that Figure 6 shows a distribution

421 pattern that similar to the urbanized areas and urban heat island effect.

422 **4.4. Limitations and future works**

423 There are still certain limitations in the present study, which could be further overcome in

424 future works. First, in this study, only moderate prediction performance is achieved in the

425 estimation of heatwave spatial patterns. The resultant model explains approximately 50-60%

426 of the spatial variability in heatwave conditions. In this study, the automatic variable selection

427 process identifies five to eleven important variable predictors in most cases. However, the

428 variable selection process only identifies three predictors in the analysis of CMA heatwave

429 events count in 2016, which are Latitude, Longitude, and LCZ_E_8000. It can be observed

430 that the inclusion of Latitude, Longitude as spatial predictors leads to the appearance of edge

431 in the prediction map of CMA heatwave events count in 2016. In that case, future work

432 should focus on further fine-tuning the model to improve the prediction performance.

433 Currently, the models are developed based on data from 86 locations (the 86 major national

434 standard meteorological stations) in the study area. The data amount is usable but not

435   sufficient for the development of very high-performance spatial models yet. The model

436   development would be beneficial from a denser spatial distribution of input data. In future

437   work, the model will be fine-tuned by using hundreds of automatic weather stations (AWS)

438   that are densely distributed in the study area. Second, in this study, so far, only the areal

439   proportion of LCZ types was calculated and used as predictors. The spatial pattern of the

440   configuration (i.e., the evenness, fragmentation, clustering, which can be quantified by means

441   of landscape ecology methods) of specific LCZ types not been analyzed yet by the present

442   study. More complex spatial pattern analysis for WUDAPT could be introduced for

443   generating more useful predictor variables in future studies to improve the model

444   performance. Last but not least, this pilot study is currently only conducted in subtropical

445   regions, which means that more tests should be conducted in other areas under different

446   climatic zones in order to verify and increase the worldwide applicability of the proposed

447   approach.

448   **5. Conclusion**

449   Investigating the spatial distribution of heatwave conditions is essential to the evaluation of

450   the heat-related vulnerability and relevant potential social and economic impacts, especially

451   for regions that are in rapid urbanization and economic development. In this study, we

452   present an effective approach to spatially estimate the heatwave patterns using machine

453   learning and WUDAPT. The approach enables direct mapping of the spatial distribution of

454   heatwave conditions by taking advantage of open urban data, which incorporates spatial

455   heterogeneity into the estimation of heatwave. The resultant models and high-resolution

456   spatial maps of heatwave generated by this study enhance the spatial assessment of heat-

457   health risk, and also provide valuable clues for how the cities should be properly planned to

458   enhance the resilience to heatwave events and heat-related disasters. The study also put up a

459   new perspective and a feasible pathway of utilizing WUDPAT to facilitate urban

460   environment applications.

461   **Acknowledgement**

467

468 **Reference**

469 Arnfield, A. J., 2003. Two decades of urban climate research: a review of turbulence,
470   exchanges of energy and water, and the urban heat island. International journal of
471   climatology. 23**,** 1-26.

472 Arnold, C. L., Gibbons, C. J., 1996. Impervious Surface Coverage: The Emergence of a Key
473   Environmental Indicator. Journal of the American Planning Association. 62**,** 243-258.

474 Bao, J., et al., 2015. The Construction and Validation of the Heat Vulnerability Index, a
475   Review. International Journal of Environmental Research and Public Health. 12**,**
476   7220-7234.

477 Bechtel, B., et al., 2015. Mapping Local Climate Zones for a Worldwide Database of the
478   Form and Function of Cities. ISPRS International Journal of Geo-Information. 4**,** 199.

479 Bechtel, B., et al., 2019. Generating WUDAPT Level 0 data – Current status of production
480   and evaluation. Urban Climate. 27**,** 24-45.

481 Bechtel, B., et al., 2016. Towards consistent mapping of urban structure - global human
482   settlement layer and local climate zones. ISPRS - International Archives of the
483   Photogrammetry, Remote Sensing and Spatial Information Sciences. XLI-B8**,** 1371-
484   1378.

485 Benmarhnia, T., et al., 2015. Review Article: Vulnerability to Heat-related Mortality.
486   Epidemiology. 26**,** 781-793.

487 Brandsma, T., Wolters, D., 2012. Measurement and Statistical Modeling of the Urban Heat
488   Island of the City of Utrecht (the Netherlands). Journal of Applied Meteorology and
489   Climatology. 51**,** 1046-1060.

490 Breiman, L., 1996. Bagging predictors. Machine Learning. 24**,** 123-140.

491 Breiman, L., 2001. Random Forests. Machine Learning. 45**,** 5-32.

492 Bunyavanich, S., et al., 2003. The Impact of Climate Change on Child Health. Ambulatory
493   Pediatrics. 3**,** 44-52.

494 Burman, P., 1989. A comparative study of ordinary cross-validation, v-fold cross-validation
495   and the repeated learning-testing methods. Biometrika. 76**,** 503-514.

496 Buscail, C., et al., 2012. Mapping heatwave health risk at the community level for public
497   health action. International Journal of Health Geographics. 11**,** 38.

498 Campbell, S., et al., 2018. Heatwave and health impact research: A global review. Health &
499   Place. 53**,** 210-218.

500   Chan, E. Y., et al., 2012. A study of intracity variation of temperature-related mortality and
501   socioeconomic status among the Chinese population in Hong Kong. J Epidemiol
502   Community Health. 66**,** 322-7.

503   Chapman, S., et al., 2017. The impact of urbanization and climate change on urban
504   temperatures: a systematic review. Landscape Ecology. 32**,** 1921-1935.

505   Chen, F., et al., 2011. The integrated WRF/urban modelling system: development, evaluation,
506   and applications to urban environmental problems. International Journal of
507   Climatology. 31**,** 273-288.

508   Ching, J., et al., 2018. WUDAPT: An Urban Weather, Climate, and Environmental Modeling
509   Infrastructure for the Anthropocene. Bulletin of the American Meteorological Society.
510   99**,** 1907-1924.

511   Ching, J., et al., WUDAPT: Facilitating advanced urban canopy modeling for weather,
512   climate and air quality applications. 2014.

513   Comrie, A. C., 2000. Mapping a Wind–Modified Urban Heat Island in Tucson, Arizona (with
514   Comments on Integrating Research and Undergraduate Learning). Bulletin of the
515   American Meteorological Society. 81**,** 2417-2431.

516   Coutts, A. M., et al., 2013. Assessing practical measures to reduce urban heat: Green and cool
517   roofs. Building and Environment. 70**,** 266-276.

518   Dugord, P.-A., et al., 2014. Land use patterns, temperature distribution, and potential heat
519   stress risk – The case study Berlin, Germany. Computers, Environment and Urban
520   Systems. 48**,** 86-98.

521   El-Zein, A., Tonmoy, F. N., 2015. Assessment of vulnerability to climate change using a
522   multi-criteria outranking approach with application to heat stress in Sydney.
523   Ecological Indicators. 48**,** 207-217.

524   Epstein, P. R., Mills, E., Climate change futures: health, ecological and economic
525   dimensions. The Center for Health and the Global Environment, Harvard Medical
526   School, 2005.

527   Field, C. B., 2012. Managing the risks of extreme events and disasters to advance climate
528   change adaptation: special report of the intergovernmental panel on climate change.
529   Cambridge University Press.

530   Fraser, A. M., et al., 2018. Strategic locating of refuges for extreme heat events (or heat
531   waves). Urban Climate. 25**,** 109-119.

532   Gál, T., et al., 2015. Comparison of two different Local Climate Zone mapping methods.

533 García-Herrera, R., et al., 2010. A Review of the European Summer Heat Wave of 2003.
534   Critical Reviews in Environmental Science and Technology. 40, 267-306.

535 Garuma, G. F., 2018. Review of urban surface parameterizations for numerical climate
536   models. Urban Climate. 24, 830-851.

537 Genuer, R., et al., 2010. Variable selection using random forests. Pattern Recognition Letters.
538   31, 2225-2236.

539 Genuer, R., et al., 2015. VSURF: an R package for variable selection using random forests.

540 Gronlund, C. J., et al., 2015. Vulnerability to extreme heat by socio-demographic
541   characteristics and area green space among the elderly in Michigan, 1990-2007.
542   Environ Res. 136, 449-61.

543 Gunawardena, K. R., et al., 2017. Utilising green and bluespace to mitigate urban heat island
544   intensity. Science of The Total Environment. 584-585, 1040-1055.

545 Guo, A., et al., 2020. Influences of urban spatial form on urban heat island effects at the
546   community level in China. Sustainable Cities and Society. 53, 101972.

547 Haines, A., et al., 2006. Climate change and human health: Impacts, vulnerability and public
548   health. Publ Health. 120.

549 Hart, M. A., Sailor, D. J., 2009. Quantifying the influence of land-use and surface
550   characteristics on spatial variability in the urban heat island. Theoretical and Applied
551   Climatology. 95, 397-406.

552 Ho, H. C., et al., 2015. A Spatial Framework to Map Heat Health Risks at Multiple Scales.
553   International Journal of Environmental Research and Public Health. 12, 16110-16123.

554 IPCC, 2014. Climate Change 2014–Impacts, Adaptation and Vulnerability: Regional
555   Aspects. Cambridge University Press, Cambridge, United Kingdom and New York,
556   NY, USA.

557 Johnson, D. P., et al., 2012. Developing an applied extreme heat vulnerability index utilizing
558   socioeconomic and environmental data. Applied Geography. 35, 23-31.

559 Johnson, S., et al., 2020. Characterization of intra-urban spatial variation in observed summer
560   ambient temperature from the New York City Community Air Survey. Urban
561   Climate. 31, 100583.

562 Kaiser, R., et al., 2007. The Effect of the 1995 Heat Wave in Chicago on All-Cause and
563   Cause-Specific Mortality. American Journal of Public Health. 97, S158-S162.

564 Kenny, G. P., et al., 2010. Heat stress in older individuals and patients with common chronic
565   diseases. Canadian Medical Association Journal. 182, 1053.

566   Klein Rosenthal, J., et al., 2014. Intra-urban vulnerability to heat-related mortality in New

567   York City, 1997–2006. Health & Place. 30**,** 45-60.

568   Kljun, N., et al., 2004. A Simple Parameterisation for Flux Footprint Predictions. Boundary-

569   Layer Meteorology. 112**,** 503-523.

570   Konarska, J., et al., 2016. Influence of vegetation and building geometry on the spatial

571   variations of air temperature and cooling rates in a high-latitude city. International

572   Journal of Climatology. 36**,** 2379-2395.

573   Kovats, R. S., Hajat, S., 2008. Heat Stress and Public Health: A Critical Review. Annual

574   Review of Public Health. 29**,** 41-55.

575   Kuhn, M., 2008. Building predictive models in R using the caret package. Journal of

576   statistical software. 28**,** 1-26.

577   Kyselý, J., 2002. Temporal fluctuations in heat waves at Prague–Klementinum, the Czech

578   Republic, from 1901–97, and their relationships to atmospheric circulation.

579   International Journal of Climatology. 22**,** 33-50.

580   Landsberg, H. E., 1981. The urban climate. Academic press, London.

581   Le Tertre, A., et al., 2006. Impact of the 2003 heatwave on all-cause mortality in 9 French

582   cities. Epidemiology. 17.

583   Leconte, F., et al., 2015. Using Local Climate Zone scheme for UHI assessment: Evaluation

584   of the method using mobile measurements. Building and Environment. 83**,** 39-49.

585   Lemonsu, A., et al., 2015. Vulnerability to heat waves: Impact of urban expansion scenarios

586   on urban heat island and heat stress in Paris (France). Urban Climate. 14**,** 586-605.

587   Li, D., Bou-Zeid, E., 2013. Synergistic Interactions between Urban Heat Islands and Heat

588   Waves: The Impact in Cities Is Larger than the Sum of Its Parts. Journal of Applied

589   Meteorology and Climatology. 52**,** 2051-2064.

590   Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news. 2**,** 18-

591   22.

592   Lin, Q., et al., 2019. The definition of heat-wave based on mortality risk assessment in

593   different regions of China. Zhonghua yu fang yi xue za zhi [Chinese journal of

594   preventive medicine]. 53**,** 97-102.

595   Liu, X., et al., 2018. High-resolution multi-temporal mapping of global urban land using

596   Landsat images based on the Google Earth Engine Platform. Remote Sensing of

597   Environment. 209**,** 227-239.

598   Luo, M., Lau, N.-C., 2016. Heat Waves in Southern China: Synoptic Behavior, Long-Term

599   Change, and Urbanization Effects. Journal of Climate. 30**,** 703-720.

Ma, W., et al., 2015. The short-term effect of heat waves on mortality and its modifiers in China: An analysis from 66 communities. Environment International. 75, 103-109.

Maragno, D., et al., 2020. Mapping Heat Stress Vulnerability and Risk Assessment at the Neighborhood Scale to Drive Urban Adaptation Planning. Sustainability. 12, 1056.

Maughan, R. J., 2012. Hydration, morbidity, and mortality in vulnerable populations. Nutrition Reviews. 70, S152-S155.

Mayrhuber, E. A.-S., et al., 2018. Vulnerability to heatwaves and implications for public health interventions – A scoping review. Environmental Research. 166, 42-54.

McCarthy, J. J., et al., 2001. Climate change 2001: impacts, adaptation, and vulnerability: contribution of Working Group II to the third assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Meehl, G. A., Tebaldi, C., 2004. More intense, more frequent, and longer lasting heat waves in the 21st century. Science. 305, 994-997.

Mills, G., et al., An Introduction to the WUDAPT project. ICUC9 - 9th International Conference on Urban Climate jointly with 12th Symposium on the Urban Environment. The International Association for Urban Climate (IAUC) and the American Meteorological Society (AMS) Toulouse, France, 2015.

Morris, K. I., et al., 2016. Effect of vegetation and waterbody on the garden city concept: An evaluation study using a newly developed city, Putrajaya, Malaysia. Computers, Environment and Urban Systems. 58, 39-51.

Nairn, J., Fawcett, R., 2011. Defining heatwaves: heatwave defined as a heat-impact event servicing all. Europe. 220, 224.

Oke, T. R., 1973. City size and the urban heat island. Atmospheric Environment (1967). 7, 769-779.

Oke, T. R., 1982. The energetic basis of the urban heat island. Quart J R Met Soc. 108.

Oke, T. R., 1987. Boundary Layer Climates Second Edition. Routledge, London and New York.

Oke, T. R., 1988. The urban energy balance. Progress in Physical geography. 12, 471-508.

Oke, T. R., 1997. Urban environments. McGill–Queen's University Press.

Oleson, K. W., et al., 2015. Interactions between urbanization, heat stress, and climate change. Climatic Change. 129, 525-541.

Oudin Åström, D., et al., 2011. Heat wave impact on morbidity and mortality in the elderly population: A review of recent studies. Maturitas. 69, 99-105.

633   Probst, P., et al., 2019. Hyperparameters and tuning strategies for random forest. WIREs Data
634   Mining and Knowledge Discovery. 9, e1301.

635   Ren, C., et al., 2020. Developing a rapid method for 3-dimensional urban morphology
636   extraction using open-source data. Sustainable Cities and Society. 53, 101962.

637   Ren, C., et al., 2019. Assessment of Local Climate Zone Classification Maps of Cities in
638   China and Feasible Refinements. Scientific Reports. 9, 18848.

639   Rey, G., et al., 2009. Heat exposure and socio-economic vulnerability as synergistic factors in
640   heat-wave-related mortality. European Journal of Epidemiology. 24, 495-502.

641   Salamanca, F., et al., 2011. A Study of the Urban Boundary Layer Using Different Urban
642   Parameterizations and High-Resolution Urban Canopy Parameters with WRF. Journal
643   of Applied Meteorology and Climatology. 50, 1107-1128.

644   Schatz, J., Kucharik, C. J., 2014. Seasonality of the Urban Heat Island Effect in Madison,
645   Wisconsin. Journal of Applied Meteorology and Climatology. 53, 2371-2386.

646   Sharma, A., et al., 2017. Urban meteorological modeling using WRF: a sensitivity study.
647   International Journal of Climatology. 37, 1885-1900.

648   Shen, J., Kee, G., 2016. Development and Planning in Seven Major Coastal Cities in
649   Southern and Eastern China. Springer International Publishing.

650   Shi, Y., et al., 2018a. Modelling the fine-scale spatiotemporal pattern of urban heat island
651   effect using land use regression approach in a megacity. Science of The Total
652   Environment. 618, 891-904.

653   Shi, Y., et al., 2018b. Evaluating the local climate zone classification in high-density
654   heterogeneous urban environment using mobile measurement. Urban Climate. 25,
655   167-186.

656   Shi, Y., et al., 2019. Assessing spatial variability of extreme hot weather conditions in Hong
657   Kong: A land use regression approach. Environmental Research. 171, 403-415.

658   Stewart, I. D., Oke, T. R., 2012. Local Climate Zones for Urban Temperature Studies.
659   Bulletin of the American Meteorological Society. 93, 1879-1900.

660   Stocker, T., 2014. Climate change 2013: the physical science basis: Working Group I
661   contribution to the Fifth assessment report of the Intergovernmental Panel on Climate
662   Change. Cambridge University Press.

663   Sun, Y., et al., 2016. Contribution of urbanization to warming in China. Nature Climate
664   Change. 6, 706-709.

665   Susca, T., et al., 2011. Positive effects of vegetation: Urban heat island and green roofs.
666   Environmental Pollution. 159, 2119-2126.

667   Taha, H., 1997. Urban climates and heat islands: albedo, evapotranspiration, and
668   anthropogenic heat. Energy and Buildings. 25**,** 99-103.

669   Taha, H., et al., 1988. Residential cooling loads and the urban heat island—the effects of
670   albedo. Building and Environment. 23**,** 271-283.

671   Tan, J., et al., 2010. The urban heat island and its impact on heat waves and human health in
672   Shanghai. International Journal of Biometeorology. 54**,** 75-84.

673   Tse, J. W. P., et al., 2018. Investigation of the meteorological effects of urbanization in recent
674   decades: A case study of major cities in Pearl River Delta. Urban Climate. 26**,** 174-
675   187.

676   Uejio, C. K., et al., 2011a. Intra-urban societal vulnerability to extreme heat: The role of heat
677   exposure and the built environment, socioeconomics, and neighborhood stability.
678   Health & Place. 17**,** 498-507.

679   Uejio, C. K., et al., 2011b. Intra-urban societal vulnerability to extreme heat: the role of heat
680   exposure and the built environment, socioeconomics, and neighborhood stability.
681   Health Place. 17**,** 498-507.

682   UN, World urbanization prospects: The 2018 revision. United Nations, Department of
683   Economic and Social Affairs (DESA), Population Division, Population Estimates and
684   Projections Section. United Nations, Department of Economic and Social Affairs
685   (DESA), Population Division, Population Estimates and Projections Section, New
686   York, 2018.

687   UN, World Population Prospects, The 2019 revision. United Nations, Department of
688   Economic and Social Affairs, Population Division, 2019.

689   Voelkel, J., Shandas, V., 2017. Towards Systematic Prediction of Urban Heat Islands:
690   Grounding Measurements, Assessing Modeling Techniques. Climate. 5**,** 41.

691   Wang, D., et al., 2019a. The impact of extremely hot weather events on all-cause mortality in
692   a highly urbanized and densely populated subtropical city: A 10-year time-series
693   study (2006–2015). Science of The Total Environment. 690**,** 923-931.

694   Wang, R., et al., 2019b. Detecting multi-temporal land cover change and land surface
695   temperature in Pearl River Delta by adopting local climate zone. Urban Climate. 28**,**
696   100455.

697   Wang, R., et al., 2018. Mapping the local climate zones of urban areas by GIS-based and
698   WUDAPT methods: A case study of Hong Kong. Urban Climate. 24**,** 567-576.

699   WMO, 2008. Guide to meteorological instruments and methods of observation (WMO-No.
700   8). Secretariat of the World Meteorological Organization, Geneva, Switzerland.

701   WMO, Weather extremes in a changing climate: Hindsight on foresight. World
702   Meteorological Organization Geneva, 2011.

703   WMO, WHO, 2015. Heatwaves and health: guidance on warning-system development.
704   World Meteorological Organization.

705   Wolf, T., McGregor, G., 2013. The development of a heat wave vulnerability index for
706   London, United Kingdom. Weather and Climate Extremes. 1**,** 59-68.

707   Wong, D. W. S., The Modifiable Areal Unit Problem (MAUP). In: D. G. Janelle, et al., Eds.),
708   WorldMinds: Geographical Perspectives on 100 Problems: Commemorating the 100th
709   Anniversary of the Association of American Geographers 1904–2004. Springer
710   Netherlands, Dordrecht, 2004, pp. 571-575.

711   Xiang, J., et al., 2013. Health Impacts of Workplace Heat Exposure: An Epidemiological
712   Review. Industrial Health. advpub.

713   Xu, Z., et al., 2012. Impact of ambient temperature on children's health: A systematic review.
714   Environmental Research. 117**,** 120-131.

715   Yu-shek, C. J., 2018. Development Of Guangdong, The: China's Economic Powerhouse.
716   World Scientific Publishing Company.

717   Zhang, Y., et al., 2017. Global climate change: impact of heat waves under different
718   definitions on daily mortality in Wuhan, China. Global Health Research and Policy. 2**,**
719   10.

720

Yuan Shi: Conceptualization, Methodology, Formal analysis, Writing - Original Draft

Chao Ren: Writing - Original Draft, Funding acquisition, Project administration

Ming Luo: Validation, Writing - Review & Editing

Jason Ching: Conceptualization

Xinwei Li: Data Curation

Muhammad Bilal: Writing - Review & Editing

Xiaoyi Fang: Resources, Data Curation, Writing - Review & Editing

Zhihua Ren: Resources