

LETTER • OPEN ACCESS

## Using clustering to understand intra-city warming in heatwaves: insights into Paris, Montreal, and Zurich

To cite this article: Yongling Zhao *et al* 2024 *Environ. Res. Lett.* **19** 064002

View the [article online](#) for updates and enhancements.

You may also like

- [Rapidly expanding lake heatwaves under climate change](#)  
R Iestyn Woolway, Eric J Anderson and Clément Albergel
- [Heatwave effects on gross primary production of northern mid-latitude ecosystems](#)  
Hang Xu, Jingfeng Xiao and Zhiqiang Zhang
- [Enhanced nighttime heatwaves over African urban clusters](#)  
Eghosa Igun, Xiyan Xu, Zitong Shi *et al.*

# Breath Biopsy Conference

BREATH BIOPSY<sup>®</sup>

Join the conference to explore the **latest challenges** and advances in **breath research**, you could even **present your latest work!**



5th & 6th November  
Online



Main talks



Early career sessions



Posters

Register now for free!

ENVIRONMENTAL RESEARCH  
LETTERS

## LETTER

Using clustering to understand intra-city warming in heatwaves:  
insights into Paris, Montreal, and Zurich

## OPEN ACCESS

RECEIVED  
5 October 2023REVISED  
15 April 2024ACCEPTED FOR PUBLICATION  
30 April 2024PUBLISHED  
10 May 2024

Original content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](#).

Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.

Yongling Zhao<sup>1,\*</sup> , Dominik Strebel<sup>1</sup> , Dominique Derome<sup>2</sup> , Igor Esau<sup>3,4</sup> , Qi Li<sup>5</sup>   
and Jan Carmeliet<sup>1</sup> <sup>1</sup> Department of Mechanical and Process Engineering, ETH Zürich, Zürich, Switzerland<sup>2</sup> Department of Civil and Building Engineering, Université de Sherbrooke, Québec, Canada<sup>3</sup> Department of Physics and Technology, The Arctic University of Norway, Tromsø, Norway<sup>4</sup> Nansen Environmental and Remote Sensing Centre, Bergen, Norway<sup>5</sup> School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, United States of America

\* Author to whom any correspondence should be addressed.

E-mail: [yozhao@ethz.ch](mailto:yozhao@ethz.ch)**Keywords:** clustering, intra-city warming, heatwaves, ground storage flux, hysteresis loop**Abstract**

We introduce a novel methodological advancement by clustering paired near-surface air temperature with the planetary boundary layer height to characterize intra-city clusters for analytics. To illustrate this approach, we analyze three heatwaves (HWs): the 2019 HW in Paris, the 2018 HW in Montreal, and the 2017 HW in Zurich. We assess cluster-based characteristics before, during, and after heatwave events. While the urban clusters identified by this clustering align well with built-up areas obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) land cover data, additional local hot spots spanning several kilometers can also be recognized, extending outside the built-up areas. Using the objective hysteresis model, we further determine the overall strength coefficient of the hysteresis loop between ground storage flux and all-wave downward radiative flux, ranging from 0.414 to 0.457 for urban clusters and from 0.126 to 0.157 for rural clusters during the heatwave periods. Across all cities, we observe a consistent refueling-restoration mode in the cumulative ground heat flux as the heatwaves progress. Future developments of this proposed two-component clustering approach, with the integration of more influential physics and advances in spatial and temporal resolutions, will offer a more comprehensive characterization of cities for urban climate analytics.

**1. Introduction**

With over half of the global population now residing in urban areas, the urban heat island (UHI) phenomenon—wherein urban areas are warmer than their surrounding rural areas (Oke 1981)—requires a profound understanding of its underlying physics (Estrada *et al* 2017). It also necessitates the development of both mitigation and adaptation strategies (Li *et al* 2014b, Krayenhoff *et al* 2018, Debnath *et al* 2023, Zhao *et al* 2023). Differences in either surface temperature (Kalnay and Cai 2003, Manoli *et al* 2020, Li *et al* 2023) or near-surface air temperature (Kubilay *et al* 2020, Chiu *et al* 2022) between urban and rural areas are commonly determined to quantify the intensity of the surface or canopy UHI (Oke 1976, 2006). It has been widely observed that the temporal variation in UHI intensity tends to be more pronounced at night

than during the day (Oke 1981, Li *et al* 2020). Insights into UHI intensity from these studies often correlate it with factors such as wind speed, anthropogenic heat emission, heat absorption by urban materials, and a variety of biophysical processes (e.g. by vegetation), especially those associated with vegetation within the urban canopy layer (Arnfield 2003, Song and Wang 2015, Wang *et al* 2018, Tan *et al* 2023).

The UHI effect emerges from the intricate interaction between surface heat fluxes and atmospheric characteristics (Kropfli and Kohn 1978, Oke 1981, Hildebrand and Ackerman 1984, Miao *et al* 2009). Above the ground, the planetary boundary layer (PBL) is the region where various forms of heat are mixed, transported, and distributed. The excessive heat from cities, a contributing factor to UHI, is retained within the atmospheric layer. Kropfli *et al* (1978) observed that roll vortices, once formed,

seemed to be locked in place in reaction to the urban environment. Hildebrand *et al* (1984) reported enhanced vertical fluxes of temperature, moisture, and along-wind stress in urban areas. They inferred that the impacts of urban surface roughness are less important compared to the influence of urban heating, especially under moderate wind conditions. A convective velocity scale, correlating the surface vertical heat flux and atmospheric turbulence, was proposed as  $w_* = \left(\frac{g}{T} \overline{T'w'} Z_i\right)^{1/3}$ , where  $\overline{T'w'}$  is the surface vertical heat flux with the prime indicating a departure from the mean value of temperature ( $T$ ) and vertical wind speed ( $w$ ),  $g$  is the gravitational acceleration, and  $Z_i$  is a representative length scale taken as the height of the inversion of the PBL above the ground (Deardorff *et al* 1982). The link between urban heat and the PBL was further observed in field measurements in smaller urban settings like Christchurch (Tapper 1990). Using data from Toulouse, Hidalgo *et al* (2008) revealed that urban breezes can transport heat from the city center's lower atmospheric to the higher atmospheric layers downwind. A recent LES-WRF study by Zhu *et al* (2016) highlighted the influence of both momentum roughness length and thermal roughness length on the surface heat flux and the PBL over urban landscapes.

During heatwave events (HWs), the PBL is influenced by both synoptic-scale atmospheric circulation and land-atmosphere interactions (Miralles *et al* 2014, Zhang *et al* 2020). During the day, the PBL height (PBLH) is notably greater during HWs compared to non-HWs, though the difference becomes less pronounced at night. The influence of high-pressure zones, combined with soil moisture-temperature feedback, has been recognized in mega-heatwaves (Miralles *et al* 2014). Furthermore, complex synergistic interactions between HWs and the UHI phenomenon have been observed (Li *et al* 2013, Zhao *et al* 2018, Kong *et al* 2023).

When the PBLH is shallow, temperature variations can be strongly amplified (Esau *et al* 2012), given the PBL's small heat capacity. The PBL inherently connects surface heat fluxes with atmospheric characteristics (Davy *et al* 2016, Lu *et al* 2023). Lin *et al* (2008) noted that anthropogenic heat, manifesting as increased surface heat flux, can elevate the PBLH by several hundred meters. Pal *et al* (2012) found that UHI intensity has a more pronounced effect on PBLH at night than during the day. Varentsov *et al* (2023) recently observed the dependence of PBLH on UHI in a sub-Arctic city. In the framework of the surface energy budget and atmospheric characteristics, Chiu *et al* (2022) proposed an analytical solution for predicting both urban heat and moisture islands.

Studies on the PBL and heat islands often begin by defining the locations of urban and rural areas, typically using land types or a city's development plan as reference points. While this method facilitates

comparative studies between urban and rural characteristics, the distinction between the two does not rely on inherent physical characteristics. Instead, it introduces subjectivity, which can sometimes obscure the definitions of urban and rural areas. This approach can be inadequate on at least two frequent occasions: (1) urban and rural classifications may not be updated promptly in rapidly developing regions, making the available surface characteristics less accurate and (2) the definitions of urban and rural areas can vary across different regions and countries. Such inconsistencies complicate the development of universal models and generalizable understanding.

We advocate for a data-driven, objective approach that relies on the intrinsic characteristics of UHI and atmospheric factors to categorize intra-city clusters for urban climate research. We first construct a paired dataset, combining two-meter air temperature and PBLH for each computational grid. This dataset then undergoes machine learning-assisted clustering to discern clusters with distinct urban or rural features. As a validation of this methodology, we applied it to three mid-latitude cities: Paris, Montreal, and Zurich, which have populations of approximately 2.2 million, 1.8 million, and 0.4 million, respectively. These cities share similar seasonal patterns and predominantly experience continental climatic effects. UHIs during three recent heatwaves in 2019, 2018, and 2017 served as the climatic conditions against which urban and rural cluster characterizations were validated. Using these identified clusters, we conducted analyses on diurnal PBL, surface heat budget, and hysteresis.

## 2. Setup of WRF simulations and two-component clustering

### 2.1. Heatwave data

We conducted mesoscale simulations of heatwaves for representative cities in periods before, during, and after the heatwaves to create spatial and temporal datasets for clustering analysis. To ensure consistency, a period of three consecutive days with peak temperatures exceeding 30 °C was classified as a heatwave. Three-day intervals closely before or after a heatwave event were selected to explore the three distinct periods. The near-surface air temperature from the simulation of each city was validated against available weather station data. The cities examined are Paris, Montreal, and Zurich. We simulated the heatwaves that occurred in 2019, 2018, and 2017 for these three cities.

### 2.2. Configuration and validation of WRF simulations

Mesoscale simulations were performed using the Advanced Research WRF (WRF-ARW, version 4.2) (Skamarock *et al* 2021). The global ERA-5 dataset from the European Center for Medium-Range Weather Forecasts was used for the initial and

boundary conditions for all the simulated cases (Hersbach *et al* 2020). The dataset, obtained from the Copernicus Climate Data service, was implemented at 3 h intervals on a  $0.25^\circ \times 0.25^\circ$  grid at both model and surface levels. The MODIS-IGBP land cover dataset, along with the GMTED2010 30-arc-second layer as the digital elevation model, were adopted for the simulations (Friedl *et al* 2002, Danielson *et al* 2011).

The resolution of the mesoscale domains was set to  $1000 \text{ m} \times 1000 \text{ m}$  for all the simulations. The spatial extent of the domains varies according to the size of the simulated cities. The latitudinal extent differed between 229 and 251 km, and longitudinally between 251 and 391 km. The number of vertical levels was set to 101 and manually configured to provide a regular spacing close to the boundary layer. No nesting was performed to save computational power, and the area of interest was placed at the center of the simulation domain. All simulations were configured with a spin-up time of 72 h before the start of the respective heatwave. The Mellor–Yamada–Janjic (MYJ) scheme was adopted for PBL parameterization given its reasonably good performance. The unified Noah land-surface model (LSM) and Eta Similarity scheme (Janjic 1994) were used in the surface layer model to implement a bulk urban parametrization, which was selected to enable primary verification of the clustering approach in the urban-rural context. The INFLUX project (Lauvaux *et al* 2016, Davis *et al* 2017, Deng *et al* 2017, Sarmiento *et al* 2017) and the studies on WRF simulation sensitivity by Li and Bou-Zeid (2014a) and Milovac *et al* (2016) suggested that the combination of MYJ scheme and Noah LSM could provide reasonably well-resolved PBL dynamics and near-surface temperatures.

To verify our WRF simulation results, near-surface 2 m air temperatures derived from WRF simulations were compared with weather station data for the three cities, as presented in table A1. The mean absolute error (MAE), root mean square error (RMSE), and  $R^2$  values suggest reasonably accurate simulations of near-surface temperatures, particularly with the incorporation of heatwave periods. The moderate discrepancy observed in the Zurich urban area is likely due to the coarse 1 km urban representation for the weather station area where the automated weather measurements could be affected by the nearby westward urban park corridor. To ensure consistent spatial resolution and urban feature representation in our clustering study for the three cities, we opted not to refine the PBL scheme and LSM combination. Instead, we applied the clustering method across the three varied cities to evaluate the framework's robustness.

### 2.3. Two-component clustering

Two-component clustering (i.e. temperature and PBLH) was performed in three steps:  $S_1$ ,  $S_2$ , and  $S_3$ , as schematically illustrated in figure 1(a). First, in

$S_1$ , we paired the air temperature  $T_2$  at 2 m with the critical PBLH on a grid level at hourly resolution, using six nocturnal time instants (1 am to 6 am) over three consecutive days during a heatwave event. This setup was chosen given the conditions of low wind speeds and a distinct urban-rural contrast (e.g. Pal *et al* 2012). The dataset is represented as  $\mathbf{D} = D_{\text{TH}(i,j,t)}$ , where  $i, j$  denote the grid index, and  $t$  denotes the time instant. Here, PBLH is estimated by scanning the local Richardson number from the ground to the height, which results in a critical Richardson number  $\text{Ri}_{\text{cr}}$  of 0.25 (Davy and Esau 2016). The corresponding PBLH was calculated using  $\text{PBLH} = g^{-1} \text{Ri}_{\text{cr}} (U^2 + V^2) T_{\text{vp}0} / (T_{\text{vp}} - T_{\text{vp}0})$ , where  $g$  is the surface gravity ( $9.807 \text{ ms}^{-2}$ );  $U$  and  $V$  represent the height-dependent local wind speed;  $T_{\text{vp}}$  is the corresponding virtual potential temperature; and  $T_{\text{vp}0}$  is the virtual potential temperature at the surface. The air temperature at 2 m ( $T_2$ ) was directly adopted from the urban parameterization simulation results.

Second, in  $S_2$ , we performed clustering analysis using Gaussian mixture models (GMM) because of their flexibility in forming clusters with unequal variances. The principle behind the application of GMM to a one-dimensional (1D) dataset is schematically illustrated in figure 1(b). This assumes that all data points are generated from a mixture of a finite number of Gaussian distributions with an unknown set of means ( $\mu$ ) and covariances ( $\sigma$ ). Similarly, clustering for the two-dimensional (2D) dataset  $\mathbf{D}$  can be performed by assuming the multivariate Gaussian distribution of  $\mathbf{D}$ , following a probability density function:

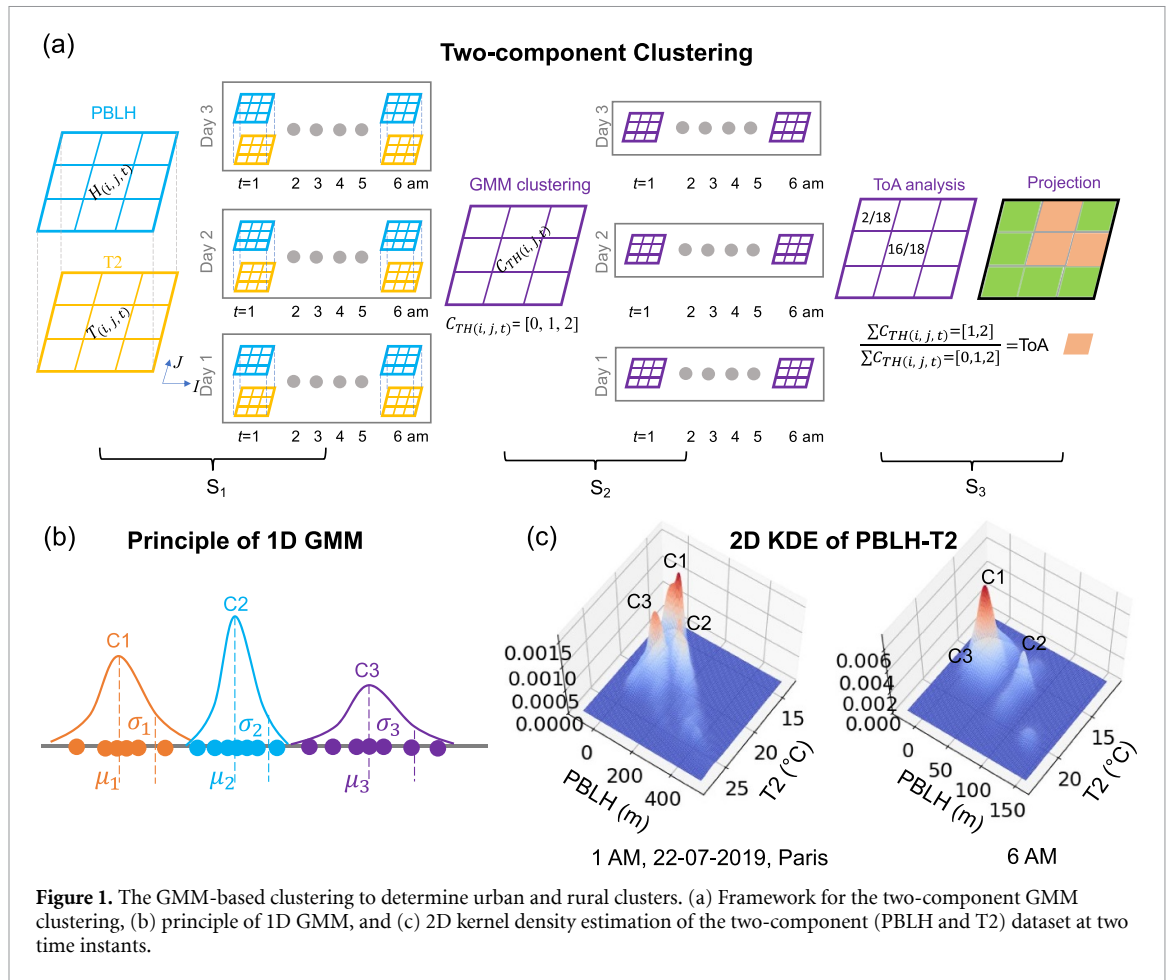
$$p(d; \mu, \sigma) = \frac{1}{2\pi^{n/2} [\sigma]^{1/2}} \exp\left(-\frac{1}{2}(d - \mu)^T \sigma^{-1} (d - \mu)\right) \quad (1)$$

where  $p(\cdot)$  denotes density functions;  $\mathbf{D} \sim \mathbb{N}(\mu, \sigma)$  has a mean matrix  $\mu \in \mathbf{M}^n$  and a covariance matrix  $\sigma \in \mathbf{CV}^n$  (Wade 2023). Examples showing the two-dimensional kernel distribution estimation of the dataset  $\mathbf{D}$  observed at two instants are presented in figure 1(c). This clearly suggests the presence of three density functions, as indicated by C1, C2 and C3.

An expectation-maximization (EM) algorithm is used to perform maximum likelihood estimation with a convergence threshold of  $10^{-3}$ . Each component is estimated with its own general covariance matrix, and the prior on weights is specified with a Dirichlet process prior. The number of clusters is chosen to be three to represent the expected rural, suburban and urban features. Depending on the specific needs and nature of the problem of interest, a different number of clusters can be used.

Third, in  $S_3$ , a ‘‘threshold-of-appearance’’ (ToA) metric was proposed and applied at the grid scale to determine whether a given grid belongs to the urban





**Figure 1.** The GMM-based clustering to determine urban and rural clusters. (a) Framework for the two-component GMM clustering, (b) principle of 1D GMM, and (c) 2D kernel density estimation of the two-component (PBLH and T2) dataset at two time instants.

cluster (i.e. C1). For each grid, the times of appearance of the clustering result being part of the urban cluster (C1), relative to all the clustering possibilities (C1, C2 and, C3), is counted as ToA:

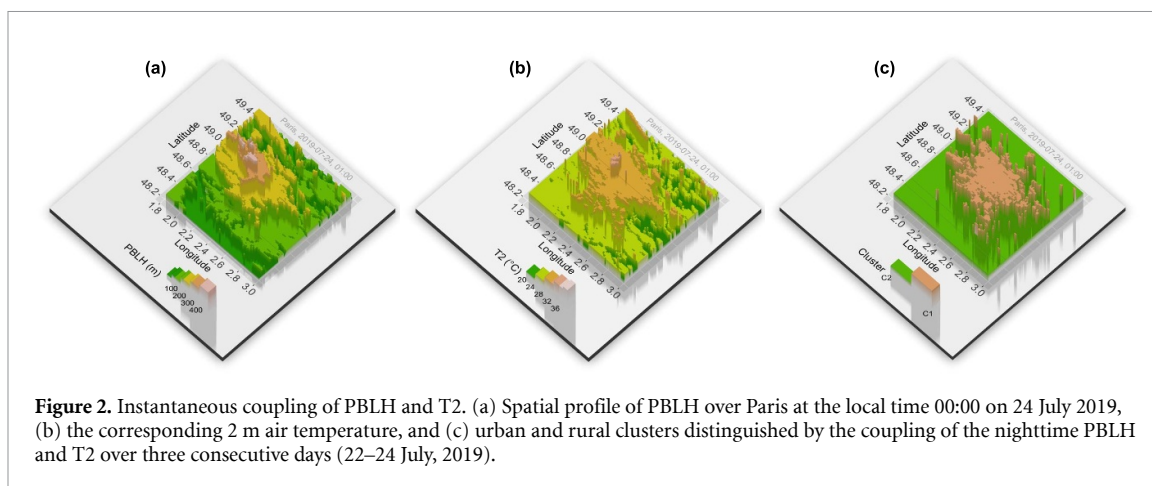
$$\text{ToA} = \frac{\sum D_{TH(i,j,t)}|C1}{\sum D_{TH(i,j,t)}|C1, C2, C3} \quad (2)$$

where C1 denotes the dataset  $D_{TH(i,j,t)}$  being recognized as a grid  $(i, j)$  of the cluster C1 at the time instant  $t$ . In our clustering analysis, the ToA in the range of 0.8 ~ 0.9 was observed to effectively identify urban areas that are in good agreement with the corresponding land type. A smaller ToA may result in an overestimated cover of urban areas, while a larger ToA usually leads to an underestimation. The use of the ToA aims to mitigate the relatively weak coupling between PBLH and T2 due to short-period strong horizontal convection during the day. Similar observations were made in the field measurements by Pal *et al* (2012), where the urban-rural contrast in PBLH was observed to occur about 85% of the time in the cycle. Zhu *et al* (2016) also noted that urban circulation diminishes when the wind speed exceeds  $5 \text{ m s}^{-1}$ . In subsequent analyses, suburban and urban clusters are agglomerated as a simplification within the scope of this study.

### 3. Results

#### 3.1. Deep heat dome above the urban cluster

In Kropfli and Kohn (1978), it was discovered that a distinct horizontal roll mode of convection was ‘locked’ to the thermal features of an urban area. This type of convection, arising from turbulent vertical heat flux, leads to a heat dome above cities (e.g. Fan *et al* 2017, Zhang *et al* 2023). Here, we initiate our analysis by examining the coupling of the PBLH and thermal features. Figure 2 presents the instantaneous spatial profile of PBLH, near-surface air temperature, and the corresponding urban cluster for Paris, as determined by the two-component clustering approach. As shown in figure 2(a), the PBLH, extending several hundred meters deep, is distinctly positioned above the Paris center. The spatial profile of the near-surface air temperature is less homogeneous, with temperatures ranging from 38 °C to 22 °C, but it still distinctly shows the Paris center (figure 2(b)). By pairing the air temperature with PBLH, the domain above the terrain can be classified into two clusters: one representing an urban cluster (in brown) and the other a rural cluster (in green), as shown in figure 2(c). These two clusters largely reflect the morphology of Paris compared to the land-use types, implying the ‘locked’ head islands



effect. Similar heat domes above the urban areas of Zurich and Montreal are also evident in our simulation results, which can be inferred in figure 3 presented in next section.

The presence of a persistent heat dome can be significantly disturbed or altered by high wind speeds. Both Hildebrand and Ackerman (1984) and Zhu *et al* (2016) noted that when the geostrophic wind speed reaches approximately  $5\text{--}6\text{ m s}^{-1}$ , the urban-centered circulation diminishes. An urban heat dome typically manifests when wind speeds are around  $2\text{ m s}^{-1}$  (Hidalgo *et al* 2008). During heatwaves, as highlighted by Li *et al* (2016), decreased wind speeds in urban areas tend to foster the formation of heat domes.

### 3.2. Cluster-based diurnal near-surface temperature and PBLH

The ‘locked’ urban clusters, along with their temperature and PBLH, are shown in figure 3. As illustrated in figures 3(a<sub>1</sub>)–(c<sub>1</sub>), two-component clustering of the PBLH and 2 m air temperature effectively distinguishes urban (in brown) and rural clusters (in green) across the three cities. This representation highlights the unique urban morphologies of each city: a centralized urban area in Paris, urban zones encircled by rivers in Montreal, and patches of urban regions adjacent to the lake in Zurich. In the clustering process, the threshold of appearance (ToA) was set at 16 out of 18, which yielded satisfactory clustering results. A similar observation was made by Pal *et al* (2012), where the consistency in PBLH variation between urban and rural areas was approximately 85%. Notably, water bodies are generally distinguishable from urban clusters. However, some exceptions at the grid level likely arise from local large-scale convective flows that alter the coupling between air temperature and the PBL.

For comparison, urban and built-up areas were also identified using MODIS-IGBP classifications, as depicted in figures 3(a<sub>2</sub>)–(c<sub>2</sub>). The urban clusters identified by clustering align well with the urban

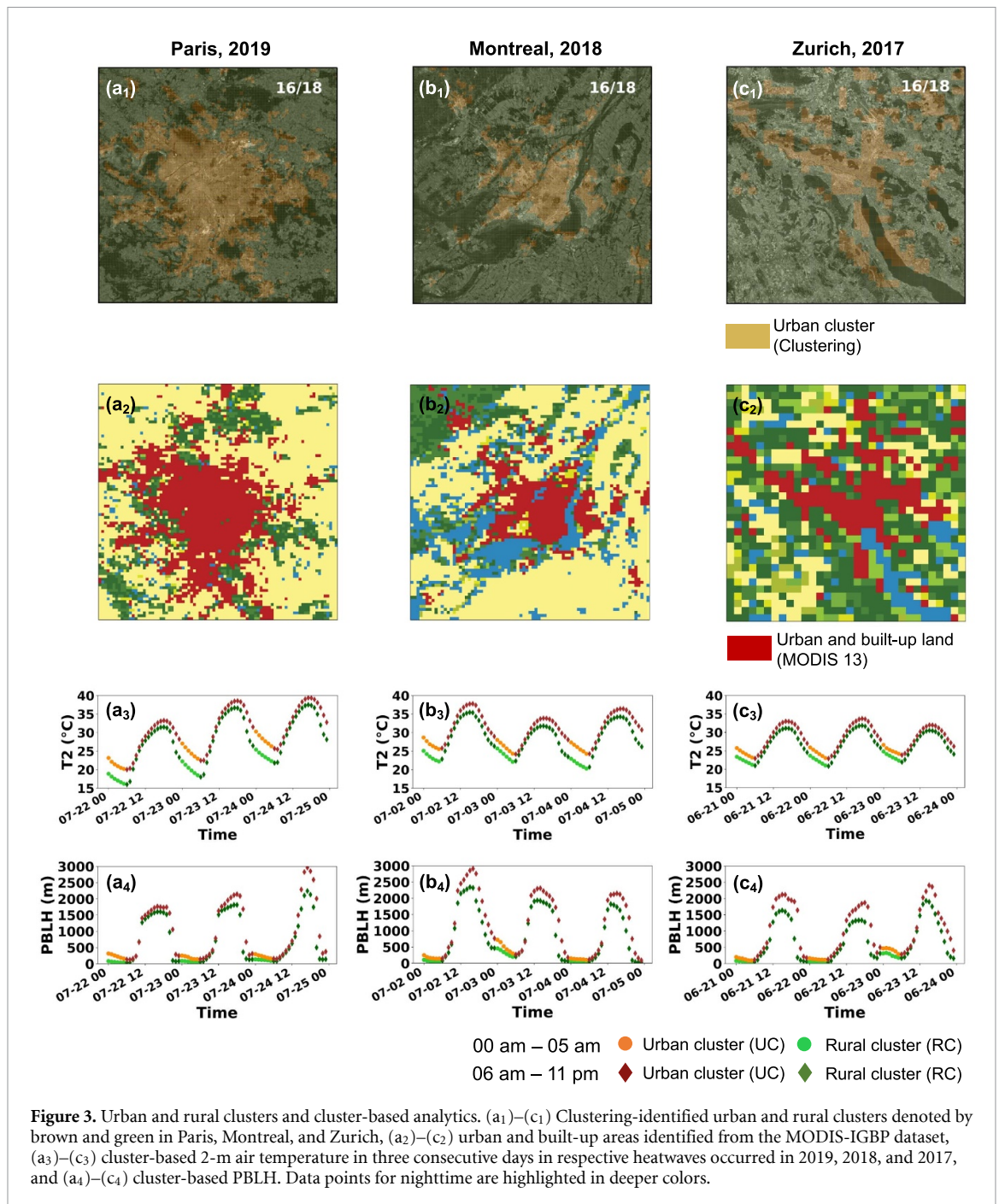
and built-up land categories obtained from MODIS-IGBP. Notably, some regions, especially in the eastern part of Paris and along Lake Zurich, extend beyond these classified urban and built-up areas. The discovery of these additional areas, which span several kilometers, suggests the presence of localized hot spots that would remain undetected with the conventional, static MODIS-IGBP data.

The clustering approach utilized to differentiate urban and rural clusters in cities offers potential for spatial ensemble analytics. First, we analyzed the ensemble averages of 2 m air temperature (figures 3(a<sub>3</sub>)–(c<sub>3</sub>)) and PBLH (figures 3(a<sub>4</sub>)–(c<sub>4</sub>)) for urban and rural clusters during heatwave days. A striking observation is the correlation between elevated air temperatures in urban clusters (orange) and an increase in PBLH.

Notably, while the air temperature difference between urban and rural clusters is more pronounced at night, PBLH differences are more distinct during the day. The nighttime air temperature difference is up to  $4\text{--}5\text{ }^{\circ}\text{C}$  between urban and rural clusters in both Paris and Montreal, whereas daytime PBLH discrepancies extend to several hundred meters. During a 4-day lidar measurement of PBLH in March 2011 under clear sky conditions in Paris, the differences in PBLH between urban and rural areas spanned from approximately  $-200\text{--}600\text{ m}$  during the day and  $0\text{--}200\text{ m}$  at night (Pal *et al* 2012). The data highlighted a robust association between PBLH and near-surface air temperature, deemed to be undeniable for both urban and suburban areas. This correlation proved to be stronger for urban dataset compared to rural ones.

### 3.3. Hysteresis of the near-surface temperature and PBLH

The observed diurnal variations in urban-rural contrasts for both air temperature and PBLH, along with the associated phase lag, suggest a hysteresis relationship between near-surface air temperature and PBLH. Here, we examine the hysteresis of the paired PBLH and 2 m air temperature for both rural and urban



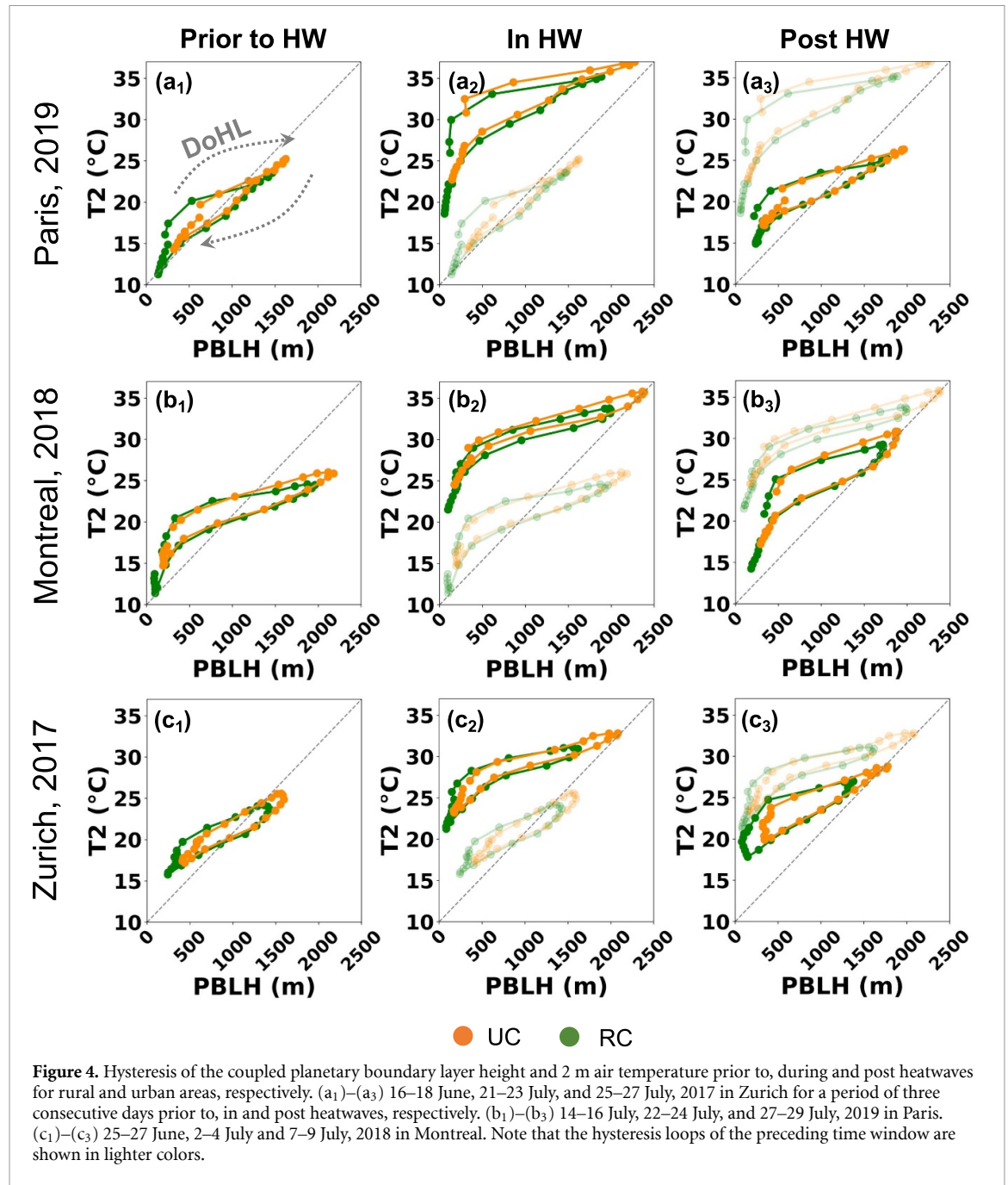
clusters during three key periods: before, during, and after heatwaves. The direction of the hysteresis loop (DoHL) is indicated. Figure 4 clearly shows that the hysteresis trajectories for the three cities vary remarkably for both rural and urban clusters during heatwaves. One of the most distinct findings is the pronounced upward expansion of the hysteresis trajectories for urban clusters on the diagram of air temperature versus PBLH. This upward shift, in stark contrast to the hysteresis loop of the preceding time window represented in lighter colors, signals a simultaneous deepening of the PBLH and an increase in air temperature during heatwave events, as shown in figures 4(a<sub>2</sub>)–(c<sub>2</sub>). Conversely, it is noteworthy that the hysteresis trajectories seen pre-heatwaves tend to

revert during post-heatwaves for both cluster types, as evidenced in figures 4(a<sub>3</sub>)–(c<sub>3</sub>). Manoli *et al* (2020) highlighted the existence of a hysteresis effect in surface UHIs on a seasonal scale, which also exhibits variability across different climatic zones. The dynamics of the hysteresis observed in the three characteristic periods remains elusive, suggesting that the underlying physics might be more intricate than solely the interplay between air temperature and PBLH.

### 3.4. Accumulative characteristics of surface heat fluxes

Surface heat fluxes are typically studied based on predefined urban and rural areas (Grimmond *et al* 1999, Hidalgo *et al* 2008). Here we analyze the surface heat





fluxes based on rural and urban clusters, and the cluster-based surface energy balance can be written as Yu *et al* (2021):

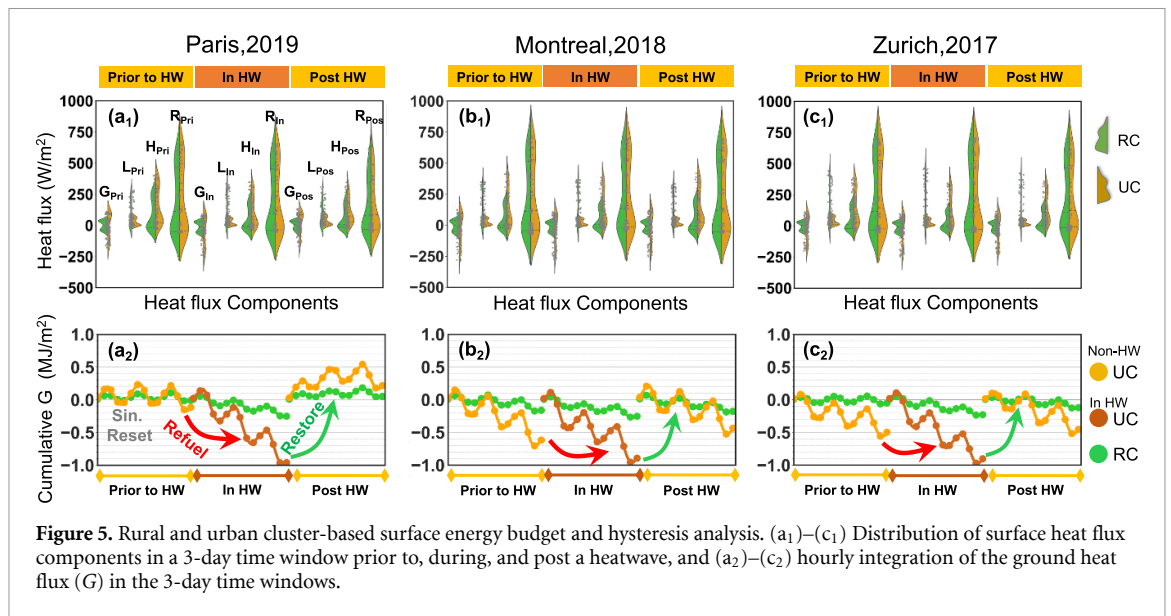
$$R + Q_{An} = H + L + G + Q_{Ad} \quad (3)$$

where  $R$  denotes the net, downward all-wave radiative flux, and  $Q_{An}$  stands for the anthropogenic heat flux.  $H$ ,  $L$  and  $G$  denote the upward surface sensible heat flux, the upward surface latent heat flux due to evaporation, and the storage heat flux of ground.  $Q_{Ad}$  denotes the heat due to advection between rural and urban clusters, which is assumed to be negligible (Chiu *et al* 2022).  $Q_{An}$  is not considered in our WRF simulations that use the Noah LSM.

Figures 5(a<sub>1</sub>)–(c<sub>1</sub>) shows the distribution of four heat flux components, resolved hourly over a 3-day period, in three cities (Paris, Montreal, and Zurich) during three distinct phases: before (denoted by subscript ‘Pri’), during (subscript ‘In’), and after (subscript ‘Pos’) a heatwave. The components include the ground heat flux ( $G$ ), latent heat flux at the surface ( $L$ ), upward sensible heat flux ( $H$ ), and incoming radiation ( $R$ ).

In all three periods, it is evident that the most significant disparity in heat flux between rural and urban clusters lies in the distribution of the ground heat flux  $G$  and latent heat flux  $L$ . The urban cluster consistently displays a broader distribution of  $G$  and a narrower band of  $L$ , implying highly inhomogeneous ground heat flux and more uniform latent heat.





**Figure 5.** Rural and urban cluster-based surface energy budget and hysteresis analysis. (a<sub>1</sub>)–(c<sub>1</sub>) Distribution of surface heat flux components in a 3-day time window prior to, during, and post a heatwave, and (a<sub>2</sub>)–(c<sub>2</sub>) hourly integration of the ground heat flux ( $G$ ) in the 3-day time windows.

This difference arises because the ground heat flux is influenced by a wider range of factors, such as surface albedo, emissivity, and urban canopy wind, compared to the factors affecting latent heat, leading to a more varied distribution of  $G$ . Similarly, a broader distribution of sensible heat flux  $H$  is observed in urban clusters across the three cities compared to their rural counterparts. As anticipated, the net radiation on the ground surface showed minimal disparity between the rural and urban clusters. Notably, the three cities exhibit similar distributions for the various surface fluxes, suggesting analogous heat transport mechanisms are at work before, during, and after the heatwaves.

However, the distribution of these 3-day heat fluxes cannot fully account for the amplified heat island intensity. A timewise integration of the ground heat flux  $G$  over their respective 3 day time windows is performed to explore possible mechanisms. This is illustrated in figures 5(a<sub>2</sub>)–(c<sub>2</sub>), where negative quantities imply downward fluxes. Before the occurrence of the heatwave in Paris, the ground heat fluxes (see figure 5(a<sub>2</sub>)) for both urban and rural clusters exhibit a clear diurnal sinusoidal pattern, denoted by ‘Sin. Reset’, allowing them to be ‘reset’ to ‘net zero.’ However, during the heatwave, this ‘reset’ cannot be maintained, leading to rapid intensification and an increase in the downward ground heat flux. After the heatwave, the diurnal features of the ground heat flux were restored. This alternating pattern in ground heat flux is significant and can be termed the refueling-restoration mode. This mode is also evident during the heatwaves observed in Montreal and Zurich (figures 5(b<sub>2</sub>) and (c<sub>2</sub>)). Notably, in both Montreal and Zurich, the intensification of the ground heat flux already starts to develop in urban clusters before the heatwaves, while the diurnal sinusoidal reset is still distinct

in their rural clusters. This early surge in heat flux may serve as a precursor to heatwaves, potentially due to soil desiccation, a phenomenon discussed by Miralles *et al* (2014). Similarly, Wang *et al* (2020) noted this distinctive transition during heatwaves in the metropolitan regions of Phoenix and New York.

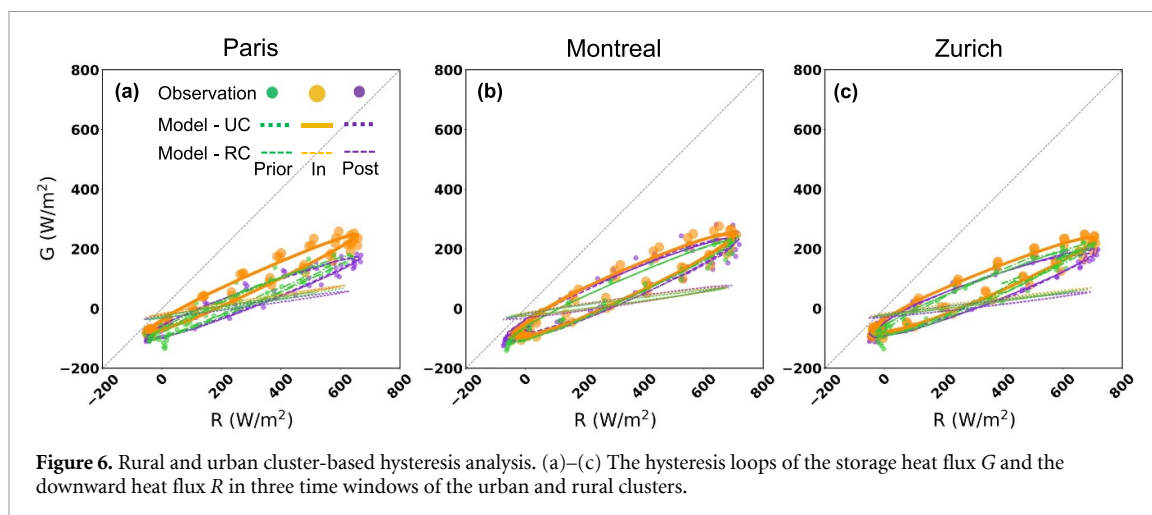
To further examine the hysteresis of the storage heat flux in relation to the all-wave downward radiative flux, the objective hysteresis model (OHM) proposed by Grimmond *et al* (1991) is adopted, which is presented for clarity:

$$G = a_1 R + a_2 \frac{\partial R}{\partial t} + a_3 \quad (4)$$

where  $a_1$  indicates the overall strength of the dependence of the storage heat flux on all-wave radiative heat flux,  $a_2$  describes the phase relations between the two fluxes, and  $a_3$  indicates the relative timing when  $G$  and  $R$  turn negative. The derivative of the all-wave downward radiative heat flux is approximated as:

$$\frac{\partial R}{\partial t} = (R_{t+1} - R_{t-1})/2. \quad (5)$$

The hysteresis loops for both urban and rural clusters during the three characteristic time windows are shown in figures 6(a)–(c). The most pronounced observation is the substantially stronger hysteresis exhibited by the urban clusters compared to the rural ones across all three cities. This observation is corroborated by the OHM coefficient  $a_1$  listed in table A2. Additional coefficients of the OHM for the heatwaves in Paris, Montreal, and Zurich can also be found in table A2. Notably, during the heatwave days, a moderate strengthening of the hysteresis is observed for the urban clusters in these cities.



#### 4. Conclusions

This study first introduces a data-driven clustering approach to characterize rural and urban clusters, leveraging the coupling of nighttime PBLH and 2 m air temperature. The clustering results from Paris, Montreal, and Zurich confirm that rural and urban regions can be distinguished well using this two-component clustering. Furthermore, this cluster-based approach facilitates deeper, physics-oriented investigations. Local hot zones, which span several kilometers and are formed by large-scale horizontal convection, are identifiable with this cluster-based approach—an analysis that is not feasible with urban boundary detection from conventional MODIS or municipal planning data.

The distinct coupling of the nighttime 2 m air temperature and PBLH—which reflects the connection between the canopy layer’s vertical heat flux and the characteristics of the PBL—can potentially be used as a two-dimensional criterion. The understanding of the transitions in the PBLH-T2 hysteresis loop across three distinct time windows could be improved with advanced hysteresis models or through analysis at a finer temporal scale. Such approaches could offer new insights for predicting the onset and decay of UHI effects during heatwaves, as illustrated using the hysteresis diagram. Additionally, an analysis of the surface energy budget based on urban and rural clusters revealed a notable intensification-restoration mode in ground storage heat flux.

While the two-component clustering method highlighted in this work effectively distinguishes intra-city heat-related phenomena, it is still primitive. Important factors such as wind speed and extensive vegetation cover are not accounted for. Specifically, when wind speeds exceed  $5 \text{ m s}^{-1}$ , this two-component approach may not distinguish between urban and rural clusters. Under such situations,

an approach employing multiple components could more effectively capture the complex interactions between ground heat and the PBL. Moreover, when using a greater number of clusters, it becomes possible to focus on the climate dynamics at the suburban or neighborhood scale. However, the impact of choosing a specific number of clusters on the results warrants careful consideration. To maximize the potential of the clustering method in studying urban climate impacts, components with higher spatial and temporal resolutions should stem from advanced data-generation simulations or be enhanced with additional meteorological measurements.

#### Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files). Data for the involved heatwave periods are available upon request.

#### Acknowledgments

Y Zhao and J Carmeliet acknowledge the SNSF project ‘Wind-driven Rain Impact of Urban Microclimate: Wetting and Drying Processes in Urban Environment’ (200021\_169323). I Esau acknowledges the support from the research and networking projects ‘URban Sustainability in Action: Multi-disciplinary Approach through Jointly Organized Research schools’ – the project of the Norwegian Research Council No. 322317; and the EEA project (Contract No. 2020TO01000219 with TACR) ‘Turbulent-resolving Urban Modelling of Air Quality and Thermal Comfort (TURBAN)’. Y Zhao would also like to acknowledge discussions with Dr. Davy Richard at the Nansen Environmental and Remote Sensing Center and Bjerknes Centre for Climate Research.

## Appendix

**Table A1.** Comparison of grid-based 2 m air temperatures between WRF-simulation results and weather station data for the full periods, including prior to, in and post heatwaves.

Cities	Rep. regions	MAE (°C)	RMSE (°C)	$R^2$	Lat (°N)	Lon (°E)
Paris	Urban (Montsouris)	1.23	1.55	0.91	48.82	2.34
	Rural (Melun)	0.89	1.18	0.96	48.61	2.68
Montreal	Urban (McTavish)	1.92	2.45	0.81	45.50	−73.57
	Rural (St-Anne)	1.81	2.37	0.83	45.43	−73.93
Zurich	Urban (Kaserne)	2.3	2.79	0.65	47.37	8.53
	Rural (Airport)	1.25	1.61	0.88	47.47	8.54

**Table A2.** OHM coefficients determined based on urban and rural clusters prior to, during and post heatwaves. The higher values of the overall strength coefficient ( $a_1$ ), which indicate stronger hysteresis in the urban cluster (UC), are highlighted in bold.

City	Cluster	Period	OHM coefficients			$R^2$
			$a_1$	$a_2$ (h)	$a_3$ ( $W m^{-2}$ )	
Paris	UC	Prior	0.370	0.254	−76.594	0.984
		In	<b>0.457</b>	0.245	−52.812	0.985
		Post	0.401	0.210	−80.293	0.955
	RC	Prior	0.135	0.050	−26.887	0.944
		In	0.157	0.032	−19.145	0.951
		Post	0.158	0.046	−28.456	0.927
Montreal	UC	Prior	0.420	0.359	−68.280	0.974
		In	<b>0.450</b>	0.370	−68.985	0.985
		Post	0.443	0.263	−83.159	0.985
	RC	Prior	0.147	0.082	−24.570	0.922
		In	0.133	0.077	−19.224	0.953
		Post	0.132	0.046	−20.275	0.957
Zurich	UC	Prior	0.361	0.333	−65.161	0.982
		In	<b>0.414</b>	0.335	−59.772	0.988
		Post	0.413	0.269	−75.682	0.974
	RC	Prior	0.116	0.055	−24.843	0.947
		In	0.126	0.044	−19.154	0.943
		Post	0.127	0.030	−22.935	0.938

## ORCID iDs


Yongling Zhao  <https://orcid.org/0000-0003-3492-0844>

Dominik Strebel  <https://orcid.org/0000-0002-3529-1798>

Dominique Derome  <https://orcid.org/0000-0002-8018-1133>

Igor Esau  <https://orcid.org/0000-0003-4122-6340>

Qi Li  <https://orcid.org/0000-0003-4435-6220>

Jan Carmeliet  <https://orcid.org/0000-0001-7489-2652>

## References

- Arnfield A J 2003 Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island *Int. J. Climatol.* **23** 1–26
- Chiu C T F, Wang K, Paschalis A, Erfani T, Peleg N, Fatichi S, Theeuwes N and Manoli G 2022 An analytical approximation of urban heat and dry islands and their impact on convection triggering *Urban Clim.* **46** 101346
- Danielson J J and Gesch D B 2011 *Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010)* (Open-File Report) (<https://doi.org/10.3133/ofr20111073>)
- Davis K J et al 2017 The Indianapolis flux experiment (INFLUX): a test-bed for developing urban greenhouse gas emission measurements *Elem. Sci. Anth.* **5** 21
- Davy R and Esau I 2016 Differences in the efficacy of climate forcings explained by variations in atmospheric boundary layer depth *Nat. Commun.* **7** 11690
- Deardorff J W and Mahrt L 1982 On the dichotomy in theoretical treatments of the atmospheric boundary layer *J. Atmos. Sci.* **39** 2096–8
- Debnath R, Bardhan R and Bell M L 2023 Lethal heatwaves are challenging India's sustainable development *PLoS Clim.* **2** e0000156
- Deng A et al 2017 Toward reduced transport errors in a high resolution urban CO<sub>2</sub> inversion system *Elem. Sci. Anth.* **5** 20
- Esau I, Davy R and Outten S 2012 Complementary explanation of temperature response in the lower atmosphere *Environ. Res. Lett.* **7** 044026

- Estrada F, Botzen W J W and Tol R S J 2017 A global economic assessment of city policies to reduce climate change impacts *Nat. Clim. Change* **7** 403–6
- Fan Y, Li Y, Bejan A, Wang Y and Yang X 2017 Horizontal extent of the urban heat dome flow *Sci. Rep.* **7** 11681
- Friedl M A et al 2002 Global land cover mapping from MODIS: algorithms and early results *Remote Sens. Environ.* **83** 287–302
- Grimmond C S B, Cleugh H A and Oke T R 1991 An objective urban heat storage model and its comparison with other schemes *Atmos. Environ. B* **25** 311–26
- Grimmond C S B and Oke T R 1999 Heat storage in urban areas: local-scale observations and evaluation of a simple model *J. Appl. Meteorol.* **38** 922–40
- Hersbach H et al 2020 The ERA5 global reanalysis *Q. J. R. Meteorol. Soc.* **146** 1999–2049
- Hidalgo J, Pigeon G and Masson V 2008 Urban-breeze circulation during the CAPITOU experiment: observational data analysis approach *Meteorol. Atmos. Phys.* **102** 223–41
- Hildebrand P H and Ackerman B 1984 Urban effects on the convective boundary layer *J. Atmos. Sci.* **41** 76–91
- Janjic Z I 1994 The step-mountain eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes *Mon. Weather Rev.* **122** 927–45
- Kalnay E and Cai M 2003 Impact of urbanization and land-use change on climate *Nature* **423** 528–31
- Kong J, Zhao Y, Strebel D, Gao K, Carmeliet J and Lei C 2023 Understanding the impact of heatwave on urban heat in greater Sydney: temporal surface energy budget change with land types *Sci. Total Environ.* **903** 166374
- Krayenhoff E S, Moustauou M, Broadbent A M, Gupta V and Georgescu M 2018 Diurnal interaction between urban expansion, climate change and adaptation in US cities *Nat. Clim. Change* **8** 1097–103
- Kropfli R A and Kohn N M 1978 Persistent horizontal rolls in the urban mixed layer as revealed by Dual-Doppler Radar *J. Appl. Meteorol. Climatol.* **17** 669–76
- Kubilay A, Allegrini J, Strebel D, Zhao Y, Derome D and Carmeliet J 2020 Advancement in urban climate modelling at local scale: urban heat Island mitigation and building cooling demand *Atmosphere* **11** 1313
- Lauvaux T et al 2016 High-resolution atmospheric inversion of urban CO<sub>2</sub> emissions during the dormant season of the Indianapolis flux experiment (INFLUX) *J. Geophys. Res. Atmos.* **121** 5213–36
- Li D and Bou-Zeid E 2013 Synergistic interactions between urban heat Islands and heat waves: the impact in cities is larger than the sum of its parts *J. Appl. Meteorol. Climatol.* **52** 2051–64
- Li D and Bou-Zeid E 2014a Quality and sensitivity of high-resolution numerical simulation of urban heat islands *Environ. Res. Lett.* **9** 055001
- Li D, Bou-Zeid E and Oppenheimer M 2014b The effectiveness of cool and green roofs as urban heat island mitigation strategies *Environ. Res. Lett.* **9** 055002
- Li D, Sun T, Liu M, Wang L and Gao Z 2016 Changes in wind speed under heat waves enhance urban heat Islands in the Beijing metropolitan area *J. Appl. Meteorol. Climatol.* **55** 2369–75
- Li X, Chakraborty T C and Wang G 2023 Comparing land surface temperature and mean radiant temperature for urban heat mapping in Philadelphia *Urban Clim.* **51** 101615
- Li Y, Schubert S, Kropp J P and Rybski D 2020 On the influence of density and morphology on the urban heat Island intensity *Nat. Commun.* **11** 2647
- Lin C-Y, Chen F, Huang J C, Chen W C, Liou Y A, Chen W N and Liu S-C 2008 Urban heat island effect and its impact on boundary layer development and land-sea circulation over northern Taiwan *Atmos. Environ.* **42** 5635–49
- Lu Y, Zhou X-H, Xiao H and Li Q 2023 Using machine learning to predict urban canopy flows for land surface modeling *Geophys. Res. Lett.* **50** e2022GL102313
- Manoli G, Fatichi S, Bou-Zeid E and Katul G G 2020 Seasonal hysteresis of surface urban heat islands *Proc. Natl Acad. Sci.* **117** 7082–9
- Miao S, Chen F, LeMone M A, Tewari M, Li Q and Wang Y 2009 An observational and modeling study of characteristics of urban heat Island and boundary layer structures in Beijing *J. Appl. Meteorol. Climatol.* **48** 484–501
- Milovac J, Warrach-Sagi K, Behrendt A, Späth F, Ingwersen J and Wulfmeyer V 2016 Investigation of PBL schemes combining the WRF model simulations with scanning water vapor differential absorption lidar measurements *J. Geophys. Res. Atmos.* **121** 624–49
- Miralles D G, Teuling A J, van Heerwaarden C C and Vilà-guerau de Arellano J 2014 Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation *Nat. Geosci.* **7** 345–9
- Oke T R 1976 The distinction between canopy and boundary-layer urban heat islands *Atmosphere* **14** 268–77
- Oke T R 1981 Canyon geometry and the nocturnal urban heat island: comparison of scale model and field observations *J. Climatol.* **1** 237–54
- Oke T R 2006 Towards better scientific communication in urban climate *Theor. Appl. Climatol.* **84** 179–90
- Pal S et al 2012 Spatio-temporal variability of the atmospheric boundary layer depth over the Paris agglomeration: an assessment of the impact of the urban heat island intensity *Atmos. Environ.* **63** 261–75
- Sarmiento D P, Davis K J, Deng A, Lauvaux T, Brewer A and Hardesty M 2017 A comprehensive assessment of land surface-atmosphere interactions in a WRF/urban modeling system for Indianapolis, IN *Elem. Sci. Anth.* **5** 23
- Skamarock W C et al 2021 A description of the advanced research WRF model version 4.3. (<https://doi.org/10.5065/1dfh-6p97>)
- Song J and Wang Z-H 2015 Impacts of mesic and xeric urban vegetation on outdoor thermal comfort and microclimate in Phoenix, AZ *Build. Environ.* **94** 558–68
- Tan H, Kotamarthi R, Wang J, Qian Y and Chakraborty T C 2023 Impact of different roofing mitigation strategies on near-surface temperature and energy consumption over the Chicago metropolitan area during a heatwave event *Sci. Total Environ.* **860** 160508
- Tapper N J 1990 Urban influences on boundary layer temperature and humidity: results from Christchurch, New Zealand *Atmos. Environ. B* **24** 19–27
- Varentsov M, Konstantinov P, Repina I, Artamonov A, Pechkin A, Soromotin A, Esau I and Baklanov A 2023 Observations of the urban boundary layer in a cold climate city *Urban Clim.* **47** 101351
- Wade S 2023 Bayesian cluster analysis *Phil. Trans. R. Soc. A* **381** 20220149
- Wang C, Wang Z-H and Sun L 2020 Early-warning signals for critical temperature transitions *Geophys. Res. Lett.* **47** e2020GL088503
- Wang C, Wang Z-H and Yang J 2018 Cooling effect of urban trees on the built environment of contiguous United States *Earth's Future* **6** 1066–81
- Yu Z, Hu L, Sun T, Albertson J and Li Q 2021 Impact of heat storage on remote-sensing based quantification of anthropogenic heat in urban environments *Remote Sens. Environ.* **262** 112520
- Zhang Y, Wang L, Santanello J A, Pan Z, Gao Z and Li D 2020 Aircraft observed diurnal variations of the planetary boundary layer under heat waves *Atmos. Res.* **235** 104801



- Zhang Y, Wang X, Fan Y, Zhao Y, Carmeliet J and Ge J 2023 Urban heat dome flow deflected by the Coriolis force *Urban Clim.* **49** 101449
- Zhao L, Oppenheimer M, Zhu Q, Baldwin J W, Ebi K L, Bou-Zeid E, Guan K and Liu X 2018 Interactions between urban heat islands and heat waves *Environ. Res. Lett.* **13** 034003
- Zhao Y *et al* 2023 Beating urban heat: multimeasure-centric solution sets and a complementary framework for decision-making *Renew. Sustain. Energy Rev.* **186** 113668
- Zhu X, Ni G, Cong Z, Sun T and Li D 2016 Impacts of surface heterogeneity on dry planetary boundary layers in an urban-rural setting *J. Geophys. Res. Atmos.* **121** 12164–79