Urbanization brings earlier onset of summertime compound heatwaves

Check for updates

Peng Ji [1,2,3], Xinyi Zhang[3], Xing Yuan [2,3,4] ✉ & Qibo Xu [5]

Urbanization is revealed to cause more frequent and hotter day-night compound heatwaves, yet its impact on their onset remains unclear. This study compares summertime compound heatwaves in 722 urban-rural station pairs globally on an event-by-event basis during 1971-2017 and attributes their difference to numerous urbanization and climatological factors using interpretable machine learning methods. Here we find a discernible earlier onset (0.23 ± 0.04 days) of compound heatwaves in cities compared to rural areas. The magnitude of this influence is primarily determined by urban building volume and height rather than the artificial impervious surface fraction, and is exacerbated by climates characterized by warm night and ample daytime solar radiation. Taller buildings significantly advance onset in cloudy, warm, and humid climates, while larger building volumes lead to earlier onset of compound heatwaves across nearly all climates. Our findings suggest that cities, particularly those with large building volumes, should issue heatwave warnings earlier.

Heatwaves are extreme climate events that pose a direct threat to human health[1,2]. Within just a few weeks, heatwaves can cause thousands of deaths[3,4], accounting for 93% of the total fatalities (~ 148,109) from hydrometeorological extreme events in Europe during 1970–2019[5]. In China, the annual death toll from heatwaves has risen from 3679 in the 1980s to 15,500 in the 2010s[6]. In addition to heatwaves occurring during the daytime, the number of days with extreme temperatures persisting day and night has significantly increased in recent decades[7–9]. These persistent day-night hot extremes, referred to as compound heatwaves (CoHot), pose greater risks and burdens on cardiovascular and respiratory health, local environments, and the communities compared to heatwaves occurring solely during the day or night[8,10].

Anthropogenic greenhouse gas emissions causing climate warming is considered the primary factor making CoHot events hotter, longer-lasting, and more frequent[8,11,12], primarily by increasing average temperatures rather than temperature variability[8]. At the local scales, urbanization can alter temperature trends[9,13], causing different changes in CoHot events between urban and rural areas. Additionally, urbanization may advance the onset date of CoHot events[14]. For instance, the 2022 heatwave in India and Pakistan[15] and the 2023 heatwave in Mediterranean region[16] occurred during atypical high-temperature seasons (e.g., spring and autumn), catching the public off-guard due to inadequate heat preparedness, thereby

exacerbating their impacts on human health and socioeconomic systems[17]. Moreover, shifts in the onset date of CoHot events exacerbates the grand challenges to early warning systems, as the predictability of heatwave onset remains limited[18]. However, compared to the extensive research on the frequency, duration, and intensity of CoHot events, current knowledge about the influence of urbanization on their onset dates remains insufficient, hindering the development of sustainable urban planning and mitigation strategies.

Unlike anthropogenic warming, the influence of urbanization on CoHot events is highly complex. For example, the contribution of urbanization to the increasing trend of CoHots in urban regions ranges from negative values to over 50%[11,13,19,20]. This highly heterogeneous effect is not only driven by varying urban fraction (usually represented by fraction of artificial impervious surface), but is also influenced by differences in the local environment background[19,21–24]. Recent studies have begun to highlight the importance of three-dimensional (3-D) urban morphology (e.g., building height, and urban volume)[25,26], but most of them focus on the urban heat island effect. The relative significance of 3-D urban morphology and urban fraction in modulating the influence on CoHots onset, remains a subject of ongoing debate[11,27].

To this end, here we identify over 700 urban–rural station pairs from more than 5000 meteorological observation stations worldwide. With

[1]State Key Laboratory of Climate System Prediction and Risk Management, Nanjing University of Information Science and Technology, Nanjing, 210044 Jiangsu, China. [2]Key Laboratory of Hydrometeorological Disaster Mechanism and Warning of Ministry of Water Resources/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, 210044 Jiangsu, China. [3]School of Hydrology and Water Resources, Nanjing University of Information Science and Technology, Nanjing, 210044 Jiangsu, China. [4]State Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, 100029 Beijing, China. [5]California NanoSystems Institute, University of California, Los Angeles (UCLA), Los Angeles, CA, 90095, USA. ✉e-mail: xyuan2@mail.iap.ac.cn

respect to these paired stations, we present a global pattern of urbanization's influences on the onset date of CoHots ($\Delta CoHot_{date}$). Employing interpretable machine learning techniques, we attribute the spatial variations in $\Delta CoHot_{date}$ to differences in environment background, 3-D urban morphology and urban fraction. The results show that the onset of CoHots in urban areas occurs, on average 0.23 days earlier than in surrounding rural areas. Sunny, warm and humid climates, along with larger urban volumes, favors earlier onsets of CoHot in cities, while tall buildings can significantly advance CoHot onset in warm and humid climates with limited solar radiation. However, no significant relationship between $\Delta CoHot_{date}$ and urban fraction is observed unless urban building volume is held constant.

Results

Impact of urbanization on the onset date of CoHot events

CoHot events are defined as periods of at least three consecutive days with extreme high temperatures during both day and night[9,13]. Urban and rural observation stations are identified based on the proportion of impervious surface area, and 722 urban–rural station pairs within 50 km of each other are selected from over 5000 stations globally (Supplementary Fig. 1 and

"Method"). For each station pair, large-scale atmospheric processes (e.g., circulation and climate change) are assumed to be identical[9], with differences in onset dates ($\Delta CoHot_{date}$) attributed to the effects of urbanization[28]. Figure 1 depicts the urban morphology characteristics and urban heat island intensity of the selected 722 urban stations. Overall, urban stations in China are located in areas with high building heights (e.g., >12 m) and high building volumes (e.g., $>2 \times 10^6$ m^3 km^2). Urban stations in the United State are primarily associated with lower building heights (e.g., 3–7 m) and building volumes ($< 12 \times 10^5$ m^3 km^2) than China, but spatial heterogeneity still exists. Urban stations in Europe (EU) capture high building height and low building volume, potentially related to a low building number density (Li et al. 2020). The average urban fraction (UrbF) during 1985–2017 does not show strong heterogeneity among China, the United States and Europe, with most values ranging from 0.3 to 0.5. Higher urban fraction, building volume, and building height are associated with lower vegetation cover (FVC), leaf area index (LAI), and albedo (BSKA and WSKA), which are likely due to the lower albedo of impervious surfaces compared to natural vegetation (Supplementary Fig. 2). Urbanization generally increases near surface air temperature, a phenomenon known of the urban heat island
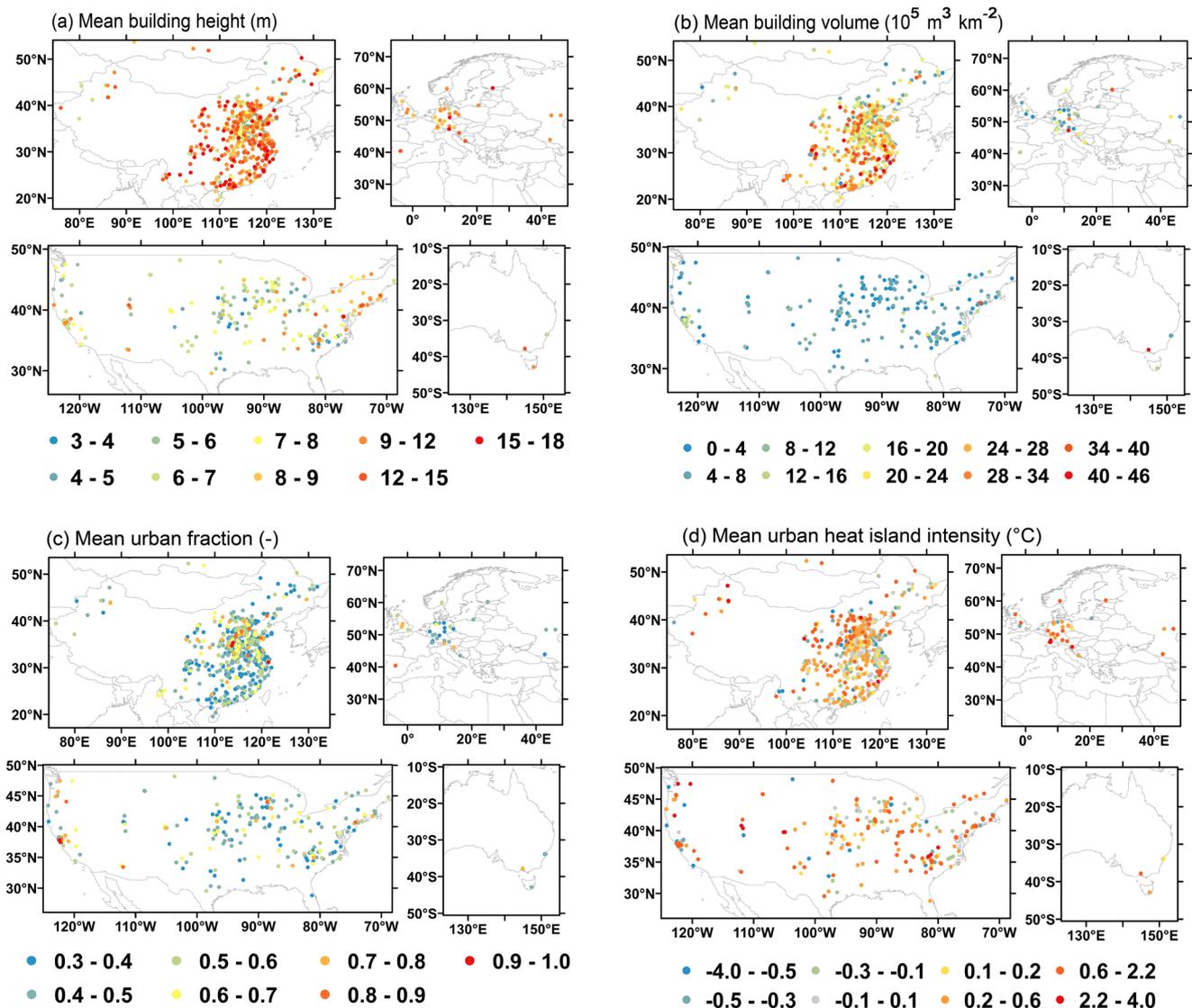

Fig. 1 | Morphological characteristics and urban heat island intensity of 722 urban stations. Spatial distributions of (**a**) mean building height, (**b**) mean building volume, (**c**) mean urban fraction and (**d**) urban heat island intensity. Urban height and volume refer to the average height of buildings and the total building volume per square kilometer, while urban fraction is the fraction of artificial impervious surface. Urban heat island intensity is the mean difference of daily maximum and minimum air temperature between the urban station and its paired rural station.

effect, that is most pronounced at night than during the day (Fig. 1d). Meanwhile, urban cooling effect can be detected over parts of northern China and western U.S., which is also found in previous studies[13,29].

Figure 2a depicts the distribution of $\Delta CoHot_{date}$ calculated through pairwise comparison of simultaneous CoHot events occurring within a 5-day window at urban-rural station pairs ("Method"). A positive $\Delta CoHot_{date}$ value indicates earlier onset of CoHot events at urban stations relative to their paired rural stations. The spatial pattern of $\Delta CoHot_{date}$ is much more heterogeneous than the distribution of urban morphology. Urbanization advances the onset of CoHot events by 0.4–4 days significantly in southern Europe, the eastern, western, and central parts of the United States, as well as along the southeastern coast and the Yangtze River Economic Belt in China (24–35°N, 100–125°E). These regions represent global hotspots of rapid urbanization, suggesting a potential link between urban development patterns and the observed advancement of CoHot event. Meanwhile, urbanization delays their onset by 0.4–0.01 days in central to northern China and midwest United States (e.g., Missouri). Overall, approximately 61% of the 722 samples (blue bars and lines in Fig. 2b), together with 81% of the 235 samples that passed the significant test, exhibits positive $\Delta CoHot_{date}$ values. When considering all samples, the observed mean $\Delta CoHot_{date}$ for China, Europe, the United States, and Australia is 0.21, 0.22, 0.26 and 0.23 days, respectively. For urban-rural pairs with significant differences in the CoHot onset dates, the regional means increased to 0.44–0.54 days (Fig. 2c). Although with some seasonal variations, the observed mean $\Delta CoHot_{date}$ keeps positive throughout the early (May-June), mid (July-August), and late (September) summer (Supplementary Fig. 3a). This pattern holds robustly across diverse climatic and geographic regions, including monsoonal versus nonmonsoonal regions that defined based on precipitation[30], and coastal versus noncoastal areas (Supplementary Fig. 3). These findings indicate that from 1971 to 2017, urbanization has generally advanced the occurrence of CoHot events on a global scale.

Subsequently, the causes of the $CoHot_{date}$ differences between urban and rural observation stations for each CoHot event are differentiated, with the results summarized in Supplementary Fig. 4. Urbanization alters the onset time of CoHot events in two main ways: (1) advancing only hot nights, and (2) advancing both hot days and hot nights. "Only hot nights" means that the onset time of hot days was the same for urban and rural stations, the onset time of hot nights is earlier at urban stations. "Both hot days and hot nights" indicates that the onset times of both hot days and hot nights are earlier at urban stations compared to rural stations. This finding is consistent with previous studies, which showed that nighttime heat extremes are more significantly influenced by urbanization, while daytime extreme heat is affected to a lesser extent[8,27,31].

Driving factors of the spatial variation in CoHot onset date due to urbanization

Previous studies have suggested that, the spatial heterogeneity of urbanization's influences on canopy temperature is modulated by numerous factors, including local climates (e.g., wind speed, humidity, and other near-surface meteorological conditions), urban fraction, 3-D urban morphology, urban greening space, and surface albedo[11,21,32,33]. We therefore employed regression methods including eXtreme Gradient Boosting (XGBoost)[34], Random Forest (RF)[35], and Multiple Linear Regression (MLR) regression methods to analyze the relationship between $\Delta CoHot_{date}$ for 722 urban–rural station pairs and various climatic and urbanization factors. Figure 3a presents the evaluation of different regression methods, with the metrics for training and validating samples shown independently in Supplementary Fig. 5. Although all methods showed a tendency to underestimate extreme positive and negative values (blue dots in Fig. 3a), XGBoost outperforms RF and MLR. Moreover, the coefficients of determination (R^2) between the simulated and observed $\Delta CoHot_{date}$ for the XGBoost, RF, and MLR models are 0.67, 0.48, and 0.02, respectively, suggesting that the XGBoost model explains the majority of the observed spatial variation in $\Delta CoHot_{date}$ (up to 67%). We also repeat the model training and evaluation using only the samples with significant $\Delta CoHot_{date}$, and the results are

consistent with those using all the samples (red dots in Fig. 3a and Supplementary Fig. 5). XGBoost's superior performance over RF is likely due to its boosting algorithm, which improves prediction accuracy by incrementally building models and focusing on correcting the errors of previous models[34].

Figure 3b provides a detailed explanation of the impact of each feature on the XGBoost model output, using SHapley Additive exPlanation (SHAP) method. Each feature, listed vertically, is associated with a distribution of SHAP values on the horizontal axis, quantifying its contribution to the model's prediction. Specifically, the sign of a SHAP value indicates the direction of the feature's influence: positive values suggest that higher feature values increase $\Delta CoHot_{date}$ (earlier onset of CoHot events in urban areas), while negative values indicate that higher feature values decrease $\Delta CoHot_{date}$. The magnitude of the SHAP value reflects the strength of the feature's influence, with larger absolute values indicating greater importance in the model's predictions. The color gradient ranges from purple to red, representing the magnitude of the feature values, with brighter colors indicating higher values. Shortwave radiation (SW) and urban volume (UrbV) demonstrate a broad range of SHAP values, indicating their significant influence on the model output across different instances. In contrast, features like mean summertime maximum temperature (TMAX) and vegetation fraction (FVC) show smaller variations in SHAP values, indicating a relatively minor impact on the prediction. In general, high shortwave radiation, high minimum temperature and high dew point temperature are associated with higher SHAP values, suggesting that sunlight-rich, warm, humid climate background is favorable to an earlier onset of CoHot events in urban regions. This is potentially because that the warm and humid climate can greatly diminish convective efficiency, leading to localized warming in urban areas[21], which in turn fosters the occurrence of heat waves. High wind speed has a negative influence on $\Delta CoHot_{date}$, as strong winds may enhance air mixing between urban and rural regions, reducing the urban heat island effect[36]. Additionally, large building volume, medium building height, and high urban fraction are conducive to the earlier onset of CoHot events. Urban engineered materials typically exhibit lower albedo and higher heat capacity than natural surfaces, enhancing solar radiation absorption. The reduction in native vegetation diminishes the evaporative cooling effect of plants and soil, causing the heat stored in the land surface to transfer into urban canopy layers primarily as sensible heat. These processes can contribute to faster warming in urban areas than rural areas[33,37], making an earlier onset of CoHot in urban areas. In summary, the data-driven XGBoost model reveals relationships between multiple factors and $\Delta CoHot_{date}$ that are physically interpretable and consistent with previous research findings.

Meanwhile, some stations exhibit delayed CoHot onset relative to their rural counterparts. This urban cooling effect aligns with previous studies documenting urbanization-induced surface and canopy temperature cooling[26,38] and reduced CoHot frequency[13]. However, the physical mechanisms underlying this cooling effect remain poorly understood[13]. Potential processes include the shading effect of buildings[38] and aerosol[39], evaporative cooling from urban water bodies and green spaces, and enhanced vertical mixing between cooler upper atmospheric air and warmer canopy air due to increased surface roughness[40]. The complex interplay of these localized processes contributes to the high spatial heterogeneity in urbanization impact. For instance, some adjacent urban stations (within 200 km) may exhibit contrasting effects on CoHot onset date (Fig. 2a). The XGBoost model generally captures these local variations (Supplementary Fig. 6a). SHAP analysis further reveals that, the opposite signs of $\Delta CoHot_{date}$ between neighboring urban stations in California and in Missouri are related to the differences in urban volume and urban greening space (Supplementary Fig. 6), respectively. Nevertheless, the XGBoost model does not fully capture all the stations with negative $\Delta CoHot_{date}$, as the model is trained globally to reveal large-scale variations of $\Delta CoHot_{date}$ and does not account for certain localized factors (e.g., aerosol and water body). Further efforts are needed to present a more in-depth investigation into the strong heterogeneity in urbanization's impact on CoHot onset dates at local scales.
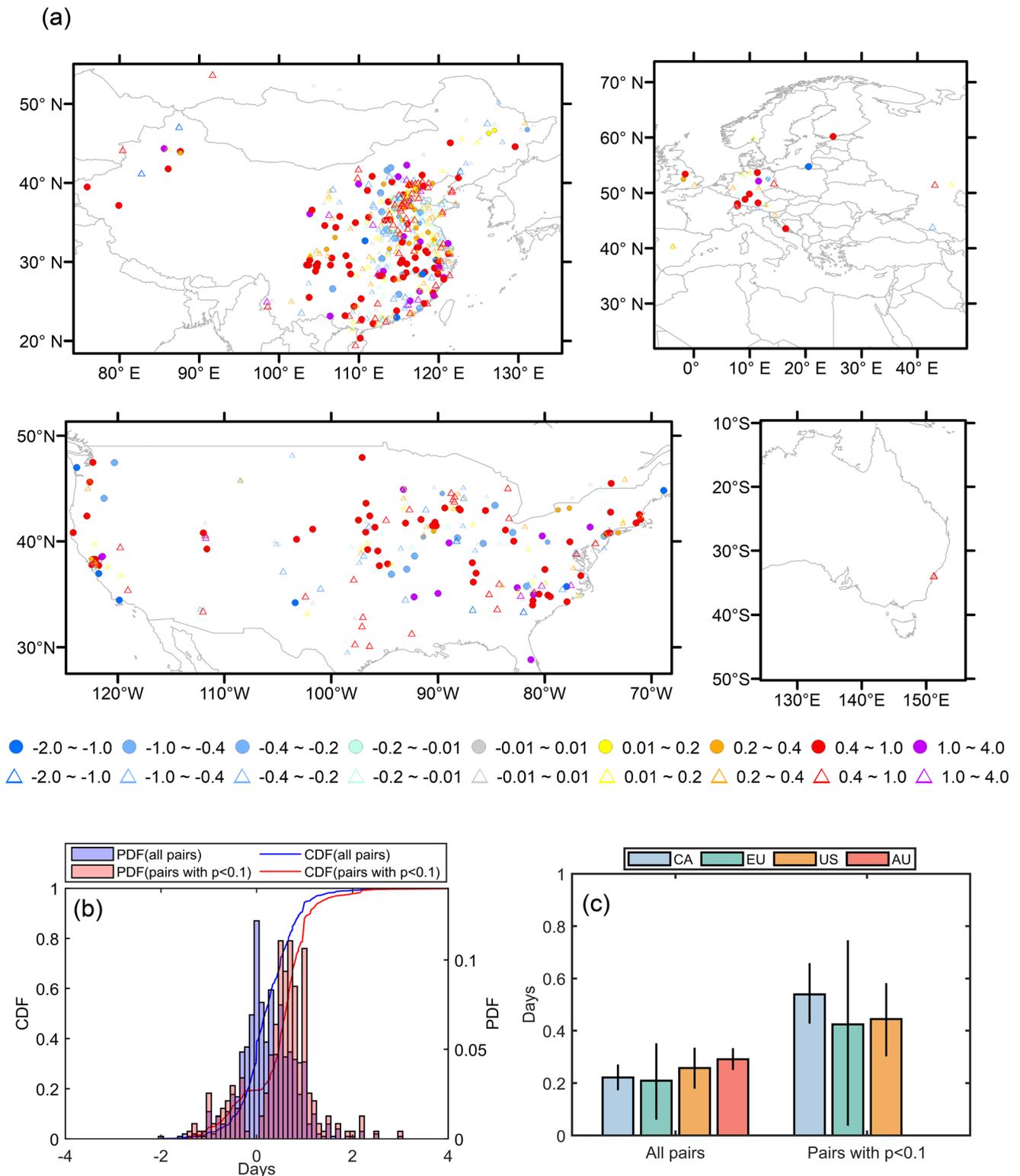
Fig. 2 | The influence of urbanization on the onset date of summertime CoHot events. a Spatial distributions of the average differences in Cohot events onset dates ($\Delta CoHot_{date}$) between urban and rural paired stations during 1971–2017. Circles mean the differences pass the student's t test at 90% confidence level. **b** Cumulative distribution function (CDF) and probability density function (PDF) of $\Delta CoHot_{date}$ from panel (**a**). **c** Mean $\Delta CoHot_{date}$ for China (CA), Europe (EU), the United State (US), and Australia (AU), with error bars representing the 5–95% uncertainty. Uncertainty estimates were derived from 10,000 bootstrap resamplings, stratified by five climate zones (Supplementary Table 1) with 80% of samples in each iteration. Note that positive $\Delta CoHot_{date}$ values indicate that CoHot event occurs earlier at urban stations than at rural stations.

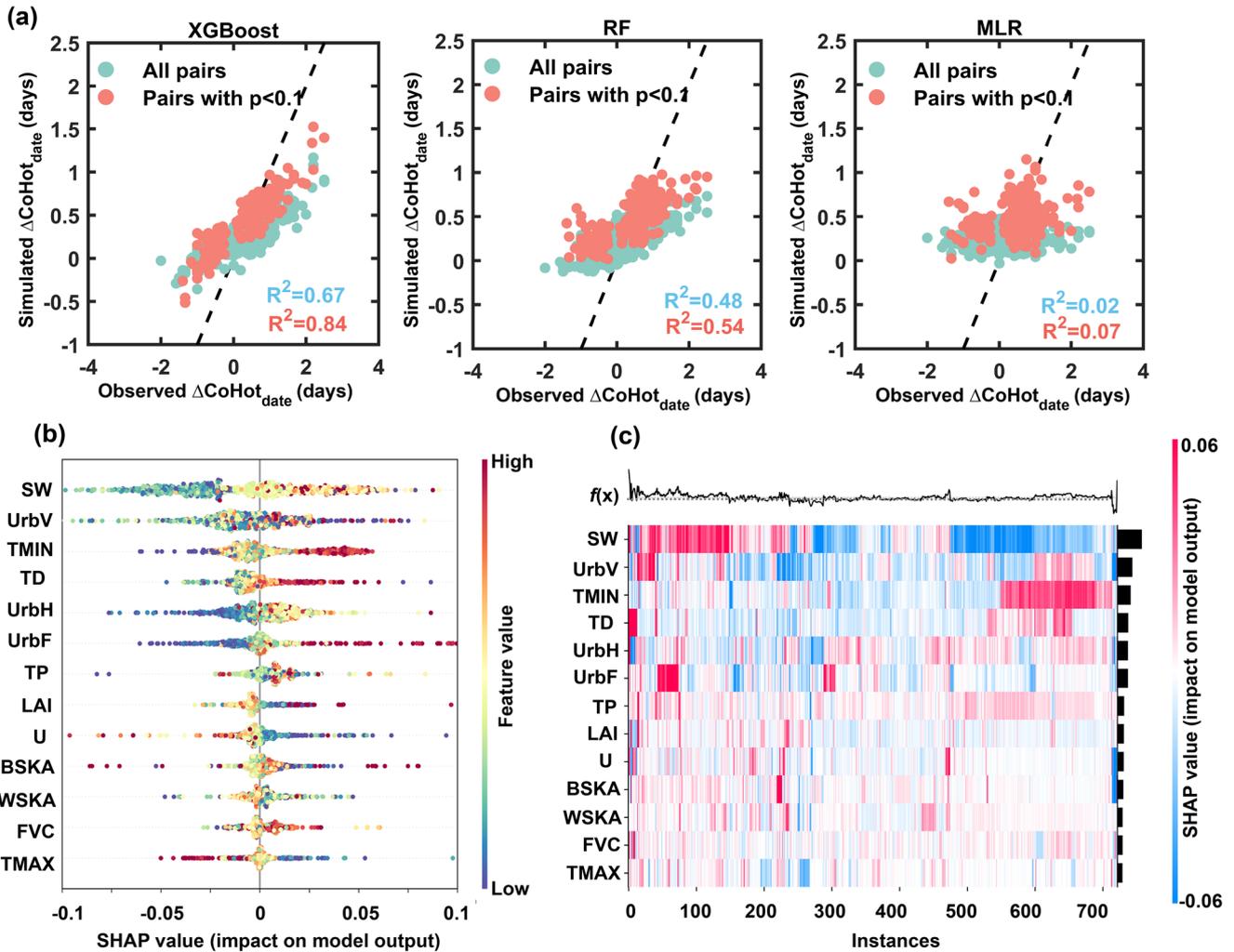**Fig. 3 | The performance of different regression models and feature importances.**
a Scatter plot of observed $\Delta CoHot_{date}$ versus the simulated values based on the XGBoost, random forest (RF), and multiple linear regression (MLR) methods.
b Beeswarm plot of SHAP values for various urban and climatic factors in the XGBoost model. Factors are arranged in descending order of the mean absolute SHAP value. **c** Heatmap of SHAP values for various factors across all model instances. The black bars on the right indicate the global importance of each factor, and the gray dashed line represents $f(x) = 0$.

The SHAP heatmap presented in the Fig. 3c illustrates the overall influence of different features on the model output. $f(x)$ represents the model's predicted value equal to the total sum of SHAP values on the vertical axis, while the black bars show the total sum of SHAP values on the horizontal axis, indicating the global importance of specific factors. Overall, shortwave radiation is the most important factor, followed by building volume, minimum temperature, building height, urban fraction, and dew point temperature. Positive $\Delta CoHot_{date}$ predictions (e.g., $f(x) > 0$) are typically associated with significant positive contributions from these factors. Other factors, such as precipitation, vegetation, wind speed, albedo, and maximum temperature, have only 1/6 to 1/3 of the importance of shortwave radiation and building volume. Therefore, the spatial variation in $\Delta CoHot_{date}$ is primarily shaped by climatic factors (e.g., solar radiation, minimum temperature, and humidity) and urban morphology.

Interdependence of urban morphology on the CoHot onset date
Considering that the impact of specific urban morphological features (e.g., building volume) on the CoHot onset date is modulated by the local climates, this study uses the K-means clustering method to categorize the 722 samples into five groups based on the top 6 environmental factors ranked in Fig. 3c. Supplementary Table 1 presents the average values of environmental factors in each of the five clusters, and the influences of environmental factors are supposed to be similar in each cluster.

Figure 4a depicts the scatter plot of urban volume and its impact on CoHot events (represented by SHAP values), with color indicating the urban fraction. Figure 4b, c display similar relationships for urban building height and urban fraction, respectively. There is a significant positive correlation between urban volume and its SHAP values (Fig. 4a). Higher building volume results in greater heat capacity (thus storing more heat), reduced wind speed, and the release of more anthropogenic heat fluxes, which enhances the urban canopy's heat effect and favors the onset of CoHot events in urban than rural regions under the same synoptic conditions (e.g., anticyclones). Meanwhile, a significant positive correlation between building height and its SHAP values is observed only in cities with high minimum temperatures, large precipitation (therefore humid climate), but low solar radiation (Fig. 4b and Supplementary Table 1). Previous studies have shown that buildings reduce urban temperatures during the day by increasing shading[25,41,42], but at night, they increase temperatures through multiple reflections and absorption of longwave radiation[26]. Therefore, in a cloudy day, warm night with humid climate background, the cooling effect of tall buildings is limited, while their warming effect at night is enhanced, potentially leading to an overall net warming effect. In regions with higher solar radiation, lower nighttime temperatures, and drier climates, the offset between the cooling and warming effects of tall buildings is more complex, and taller buildings do not always lead to an earlier CoHot onset date.
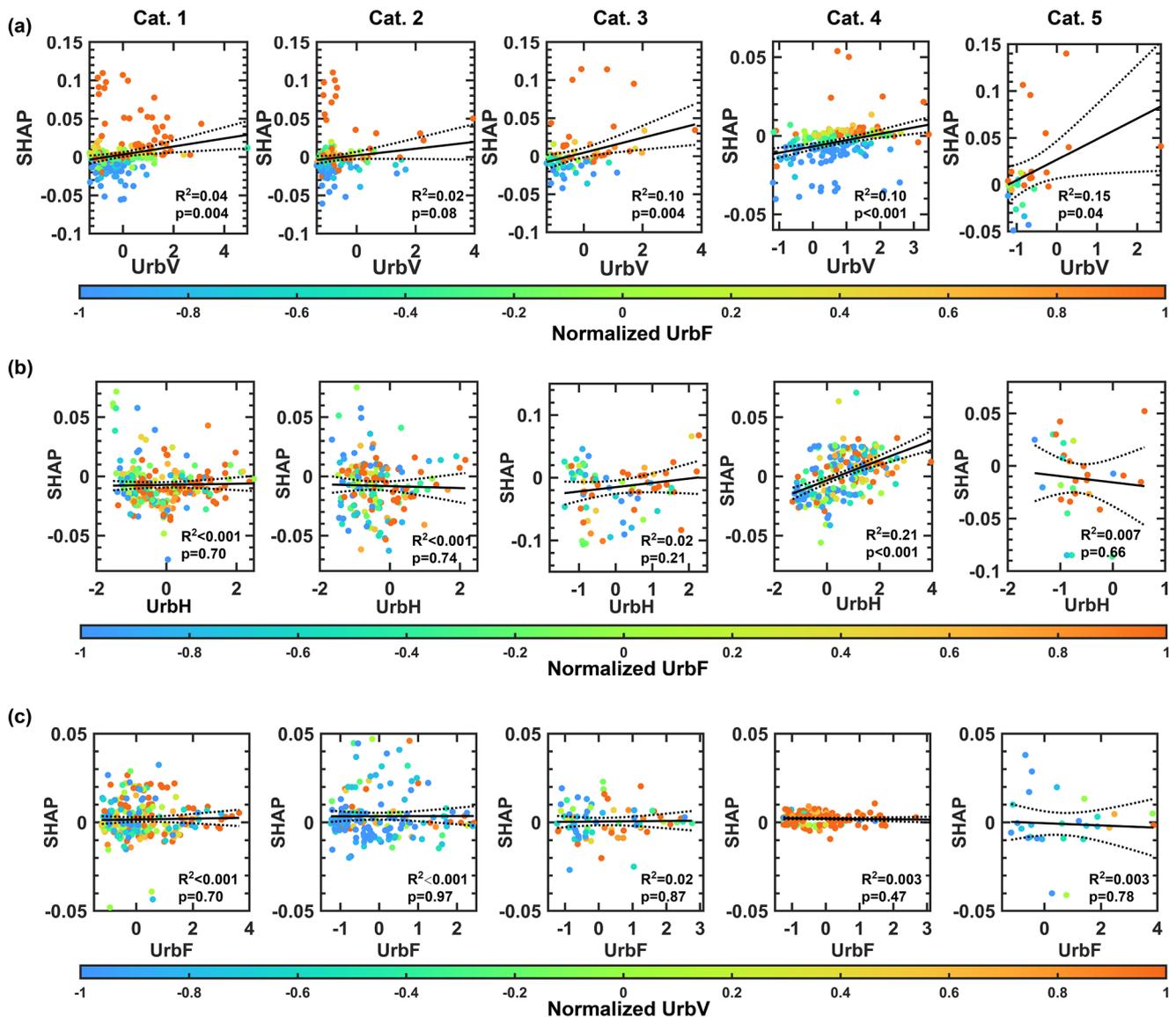
Fig. 4 | The dependency between urban factors and their impacts on $\Delta CoHot_{date}$.
a Scatter plot of the normalized urban volume (UrbV) versus its impacts on
$\Delta CoHot_{date}$ (represented by SHAP value), with colors indicating normalized urban
fraction (UrbF). The five clustered climatic categories (e.g., from Category 1 to
Category 5) are detailed in Supplementary Table 1. The black solid and dashed lines
represent linear regression and its 5–95% uncertainty, respectively. Panels (**b**) and
(**c**) Same as (**a**), but showing the influence of urban height (UrbH) and urban fraction
(UrbF), respectively. The normalization was performed using z-score normalization:
$(X - \overline{X})/\sigma$, where X means a specific variable (e.g., urban fraction), and $\overline{X}$ and σ are
the mean value and standard deviation of X from all urban stations.

There is no significant positive correlation between the urban fraction
and its SHAP values (Fig. 4c), but a higher urban fraction leads to significantly
higher positive SHAP values when urban volume density if fixed (blue to
orange dots in Fig. 4a). This suggests that the influence of urban fraction on
$CoHot_{date}$ depends on urban volume density. In contrast, the effects of urban
volume density do not depend on the urban fraction (Fig. 4c). The observed
interdependence is not attributable to collinearity between urban fraction and
urban volume density (correlation coefficient of 0.23), because (1) the
XGBoost model utilizes regularization techniques to address multi-
collinearity, and (2) there is no significant dependence of urban fraction on
urban building height (Fig. 4b), despite the high correlation between urban
building height and urban volume density. Instead, these findings highlight
the greater importance of urban volume density over urban fraction in
determining the impact of urbanization on CoHot onset.

Discussion
Compared to previous studies focused on the frequency, duration, and
intensity of CoHot events, this research goes a step further by investigating
the impacts of urbanization on the onset of CoHot events, which is
important for heatwave early warning. The result shows a global emerging of
earlier CoHot event onset in urban regions, and the sunny, warm, and
humid climate background can facilitate this advancement. In addition,
urban volume has the strongest positive influence on $\Delta CoHot_{date}$, compared
to urban building height and urban impervious fraction. Higher building
height significantly increases $\Delta CoHot_{date}$ only in the cloudy day, warming
night and humid climate, while urban impervious fractions show no sig-
nificant relationship with $\Delta CoHot_{date}$.

While the mean urban-rural differences in CoHot onset
(0.21–0.23 days globally, 0.44–0.54 days for specific locations) may appear
modest at first glance, they are statistically significant at both global and
regional scales, demonstrating a systematic impact of urbanization.
Importantly, there is considerable variability in individual events, as not all
CoHot events during 1971–2017 were advanced by urbanization. For
instance, when synoptic system anomalies are strong, CoHot events in
rural-urban station pairs may occur on the same day, accounting for 53% of
total CoHot events (Supplementary Fig. 7a). However, in the remaining

cases where urbanization drives an earlier onset (35% of events), the advance is typically 1–2 days (Supplementary Fig. 7b and Supplementary Fig. 8), exceeding the operational timescales of most heatwave warning systems (e.g., 1 day). Moreover, the proportion of advanced CoHot events exceeds 50% of total events in regions including southern Europe, the southeastern and western United States, as well as along China's southeastern coast and the Yangtze River Economic Belt (Supplementary Fig. 7b). As global urbanization intensifies and compound heatwave events become more frequent and severe in the future, the impact of urbanization on CoHot onset is likely to grow. Therefore, while synoptic-scale anomalies may sometimes synchronize CoHot onset across urban-rural pairs, incorporating urbanization-induced timing differences into warning systems is critical for enhancing urban preparedness and targeted interventions in vulnerable regions such as southeastern China, the southeastern and western U.S., and southern Europe. Additionally, our findings on the role of urban three-dimensional structure suggest that reducing urban building volumes, alongside traditional mitigation strategies like green roofs and urban greening, could effectively mitigate an earlier CoHot onset. Sustainable urban development strategies could consider reducing urban building volumes while meeting urbanization demands, thereby reducing the risk of an earlier CoHot onset.

Previous studies defined the onset date of CoHot as the time when the first CoHot event occurs in a particular summertime of a given year[14,43]. However, this definition is more applicable to phenomena with average state variables or long-term events (e.g., monsoons), rather than extreme events. This study focuses on CoHot events occurring simultaneously in both urban and rural stations, comparing each event individually. The selected sample represents about half of all CoHot events in global urban stations (Supplementary Fig. 9), while the remaining samples, which only occur in urban stations, are considered to reflect the influence of urbanization on CoHot frequency, possibly due to highly localized processes or observational perturbations[14]. The spatial threshold used to identify urban-rural station pairs should neither be too small nor too large to ensure that stations are paired under homogeneous synoptic conditions while including as many pairs as possible. Therefore, a 50 km threshold was chosen, as it can cover most medium to large-scale synoptic systems that are typically associated with heatwave extremes[44–47], and the threshold falls within the practical ranges (20–100 km) commonly used in previous studies[48,49]. Moreover, the results are not sensitive to the spatial threshold, as the conclusions remain unchanged when the threshold changes between 10 km and 100 km (Supplementary Fig. 10).

Due to the lack of long-term observation stations, the current urban-rural pairs are mainly concentrated in China, United States, and Europe, with large cities not fully represented. For example, the average building volume in U.S. is larger than in China[50], but most of U.S. urban stations identified in this study have lower building volumes than those in China. Therefore, the average earlier start date for urban CoHot events may be underestimated. Satellite-based spatial grid datasets for daily minimum and maximum surface air temperature could provide a more comprehensive urbanization impact pattern[51], but how to apply satellite short-term records to extreme event analysis (e.g., CoHot) still requires further exploration.

While the XGBoost model demonstrates superior performance, some uncertainties remain. First, the reliance on static urban 3-D morphology data may not adequately capture the dynamic changes during urban development, limiting the model's ability to accurately reflect the complex relationships between climatic and urbanization factors. In addition, neglecting this dynamic process would underestimate the importance of urban height and volume in influencing the spatial variations of $\Delta CoHot_{date}$. Second, while our predictor variables were carefully selected, they may not fully capture all relevant processes, potentially introducing biases in the predictions. Specifically, although we primarily attribute urban-rural differences in CoHot onset to urbanization effects, other small-to-meso scale (≤ 50 km) processes—such as regional topography and associated climate variability[52], and land-sea contrasts[53],—may also contribute to the differences. These factors can interact with urbanization processes, modifying

local climate conditions and heatwave dynamics. However, it is important to note that dynamic urban morphology data are currently lacking, and the factors influencing urban CoHot events, along with their underlying physical processes, remain an active area of research. While these uncertainties need further investigation, they are not expected to significantly impact the validity of our current findings, as the XGBoost model accounts for the majority of the spatial variance in CoHot onset and produces physically reasonable results.

Finally, while our machine learning framework (XGBoost with SHAP analysis) robustly identifies urban morphological features as key predictors of $\Delta CoHot_{date}$—demonstrating consistent, physically plausible relationships – we recognize that data-driven approaches alone cannot fully elucidate the underlying physical mechanisms. The relative contributions of processes like heat storage, roughness effects, and radiation trapping in driving earlier CoHot onset remain to be quantitatively resolved through process-based modeling based on urbanized climate models (e.g., WRF-UCM) or high-resolution large eddy simulations (LES). We also acknowledge that energy use patterns and infrastructure systems (e.g., air conditioning prevalence) can modify urban thermal environments, but these dynamic variables present measurement challenges for comparative analysis across diverse cities due to data availability. Future research could productively build on our findings that mainly focus on urban morphological foundation by incorporating dynamic human-system variables where reliable data become available, particularly through city-specific case studies and event-based dynamic modelings.

Methods
Data
The daily maximum temperature (TMAX) and minimum temperature (TMIN) data for the summer months (May-September and November-March in Northern and Southern Hemispheres, respectively) from 1971 to 2017 were obtained from China Meteorological Administration (CMA) and the Global Historical Climatology Network-Daily (GHCND) database[11,54]. If there were 5 or more consecutive days of missing data during a summer, that season was considered a missing summer, and stations with 5 or more missing summers were excluded from the analysis[11]. There are no differences between GHCND and CMA datasets over China, except that the latter provides observations from more stations. The CMA dataset is necessary to ensure adequate spatial coverage in China, as the GHCND dataset contains only 177 stations over China after quality control, which is substantially fewer than the >2000 stations available in the U.S. Moreover, no urban-rural pairs can be found from GHCND dataset in China.

Urbanization data from 1985 to 2017 were derived from the 30-meter global artificial impervious areas (GAIA)[55]. The urbanization percentage for each station was derived by calculating the proportion of artificial impervious area within a 2-km radius buffer zone around the station[8]. Urban morphology data, including building height and volume, were obtained from the 1-km resolution global three-dimension building structure dataset[50]. Here the urban morphology was assumed to be constant due to a lack of global dataset of dynamic 3-D urban building structure. Climate background data during the summertime (1971–2017), including downward shortwave radiation (SW), longwave radiation (LW), dew point temperature (TD), total precipitation (TP), and 10-meter wind speed (U), were sourced from the European Center for Medium-Range Weather Forecasts (ECMWF) Fifth Generation Atmospheric Reanalysis (ERA5)[56]. Vegetation and albedo data, including leaf area index (LAI), fraction of vegetation cover (FVC), white sky albedo (WSA), and black sky albedo (BSA), were derived from the Global LAnd Surface Satellite (GLASS) dataset[57]. The spatial resolution of the ERA5 and GLASS datasets were 0.25° and 0.05°, respectively. The nearest neighbor interpolation method was used to map these gridded data to each urban station.

Definitions of CoHot events and the influence of urbanization
Similar to previous studies[8,11,13], CoHot events are defined as extreme high-temperature events lasting at least three consecutive days, during which both

the daytime maximum temperature and nighttime minimum temperature exceed their respective 90th percentile thresholds. The daily 90th percentile thresholds for maximum (TMAX) and minimum (TMIN) temperatures were estimated based on a 15-day window over the baseline period (1971–2000) using 450 samples (e.g., 15 days × 30 years). These daily-based percentile thresholds, in contrast to seasonal-fixed thresholds, incorporate the varying levels of preparedness and acclimatization of both humans and ecosystems to extreme heat throughout the season[8]. In addition, utilizing a 15-day moving window helps mitigate potential inconsistencies in the percentile-based indices of temperature extremes[8].

A station is considered urban area if its average UrbF during 1985–2017 exceeds 0.33. Other stations are categorized as rural area[9]. Urban stations are paired with rural stations within a 50-km radius, assuming that both stations experience similar large-scale atmospheric processes (e.g., atmospheric circulation and climate change). The $\Delta CoHot_{date}$ between urban and rural stations was calculated through event-by-event comparison. Specifically, if a CoHot event occurs at a particular urban station, and within a 5-day window at its paired rural station, the CoHot events at both urban and rural stations are considered as the same event. The difference in their start dates (e.g., $\Delta CoHot_{date}$) is then calculated. However, if the CoHot event at the urban station is unrelated to the event at the rural station (e.g., due to the influence of urbanization on CoHot frequency), that event is excluded from the analysis. The mean $\Delta CoHot_{date}$ across all rural-urban paired stations is taken as the urbanization effect on the CoHot onset date. The student's t test is used to determine whether the average $\Delta CoHot_{date}$ is significantly positive or negative. It is important to note that CoHot events at urban stations were excluded when the annual urban fraction (UrbF) ≤ 0.33, typically reflecting early urbanization stages. Similarly, events at rural stations were excluded when UrbF > 0.33, a condition often observed in later years as rural areas urbanize. This ensures all samples reflect clear urban-rural contrasts.

Regression models and Shapley additive explanation (SHAP) method

This study employed Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Multiple Linear Regression (MLR) methods. Random Forest is a classical ensemble machine learning method that constructs multiple decision trees randomly and uses ensemble predictions to improve accuracy and stability[35]. XGBoost is an improved implementation of Gradient Boosting Decision Trees (GBDT). Unlike Random Forest, XGBoost generates multiple decision trees iteratively, where each new tree fits the errors made by the previous tree[34]. XGBoost also employs regularization techniques that help mitigate the impact of multicollinearity by penalizing overly complex models. This regularization not only reduces overfitting but also allows the model to identify and weigh the contributions of each feature more accurately, even in the presence of correlated variables. In addition, the tree-based nature of XGBoost allows it to capture complex interactions between features. Each decision tree in the ensemble can split based on different variables, effectively learning how the relationships between features influence the target variable. This structure helps the model to discern the independent effects of each feature while considering their interactions, therefore reduces potential confounding effects between different variables. In recent years, both Random Forest and XGBoost have been widely used in regression analysis, demonstrating high modeling accuracy.

Above regression models were trained and evaluated based on the cross-validation method. Specifically, the $\Delta CoHot_{date}$ of all urban-rural pairs was randomly divided into two groups, with 622 pairs (185 pairs for samples with significant $\Delta CoHot_{date}$ changes) utilized for training and 100 (50 pairs for samples with significant $\Delta CoHot_{date}$ changes) reserved for validation. Hyperparameter optimization was performed, encompassing key parameters such as the learning rate, number of estimators (n_estimators), maximum tree depth (max_depth), and gamma factors, utilizing the GridSearchCV module in Python.

Shapley additive explanation (SHAP) is a game-theory-based method for explaining machine learning model output[58]. SHAP quantifies the specific contribution of each input element to the prediction outcome. Its additive attribution formula is as follows:

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^{N} \phi_i x' \tag{1}$$

where $f(x)$ is the original model, $g(x')$ is the explanatory model with simple input x' ($x' \in \{0, 1\}^N$), while x and x' are linked by the mapping equation $x = h_x(x')$. The total number of the model input predictors is denoted by N, while ϕ_i is the feature attribution function of the ith predictor. The explanatory model $g(x')$ has a unique solution:

$$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!(M - |z'| - 1)!}{M!} (f_x(z') - f_x(z'i)) \tag{2}$$

where $|z'|$ is the number of non-zero values in z', and $z' \subseteq x'$, $f(x') = f(h_x(z')) = E[f(z)|z_S]$, and S is the set of non-zero indexes in z'.

Data availability

Code availability

References
1. Schär, C. CLIMATE EXTREMES The worst heat waves to come. *Nat. Clim. Change* **6**, 128–129 (2016).
2. Lüthi, S. et al. Rapid increase in the risk of heat-related mortality. *Nat. Commun* **14**, https://doi.org/10.1038/s41467-023-40599-x (2023).
3. Fouillet, A., et al. Excess mortality related to the August 2003 heat wave in France. *Int Arch. Occ. Environ Health* **80**, 16–24 (2006).
4. Robine, J. M. et al. Death toll exceeded 70,000 in Europe during the summer of 2003. *Cr Biol.* **331**, 171–U175 (2008).
5. World Meteorological Organization (WMO). Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970–2019). WMO-No. 1267 https://library.wmo.int/viewer/57564/download?file=1267_Atlas_of_Mortality_en.pdf&type=pdf&navigator=1 (2021).
6. Chen, H. Q. et al. Spatiotemporal variation of mortality burden attributable to heatwaves in China, 1979-2020. *Sci. Bull.* **67**, 1340–1344 (2022).
7. Chen, Y. & Zhai, P. M. Revisiting summertime hot extremes in China during 1961-2015: Overlooked compound extremes and significant changes. *Geophys Res Lett.* **44**, 5096–5103 (2017).
8. Wang, J. et al. Anthropogenically-driven increases in the risks of summertime compound hot extremes. *Nat. Commun* **11**, https://doi.org/10.1038/s41467-019-14233-8 (2020).
9. Wang, J. et al. Anthropogenic emissions and urbanization increase risk of compound hot extremes in cities. *Nat. Clim. Change* **11**, 1084–1089 (2021).

10. Liu, J. D. et al. Rising cause-specific mortality risk and burden of compound heatwaves amid climate change. *Nat. Clim. Change*, https://doi.org/10.1038/s41558-024-02137-5 (2024).

11. Ji, P., Yuan, X., Ma, F. & Xu, Q. B. Drivers of long-term changes in summer compound hot extremes in China: Climate change, urbanization, and vegetation greening. *Atmos. Res.* **310**, https://doi.org/10.1016/j.atmosres.2024.107632 (2024).

12. Wang, X. X., Lang, X. M. & Jiang, D. Detectable anthropogenic influence on summer compound hot events over China from 1965 to 2014. *Environ. Res. Lett.* **17**, https://doi.org/10.1088/1748-9326/ac4d4e (2022).

13. Ma, F. & Yuan, X. More Persistent Summer Compound Hot Extremes Caused by Global Urbanization. *Geophys. Res. Lett.* **48**, https://doi.org/10.1029/2021GL093721 (2021).

14. Igun, E., Xu, X., Shi, Z. & Jia, G. Enhanced nighttime heatwaves over African urban clusters. *Environ. Res. Lett.* **18**, https://doi.org/10.1088/1748-9326/aca920 (2022).

15. Zachariah, M. et al. Attribution of 2022 early-spring heatwave in India and Pakistan to climate change: Lessons in assessing vulnerability and preparedness in reducing impacts. *Environ. Res. Clim.* **2**, https://doi.org/10.1088/2752-5295/acf4b6 (2023).

16. Lemus-Canovas, M., Insua-Costa, D., Trigo, R. M. & Miralles, D. G. Record-shattering 2023 Spring heatwave in western Mediterranean amplified by long-term drought. *Npj Clim. Atmos. Sci.* **7**, https://doi.org/10.1038/s41612-024-00569-6 (2024).

17. Campbell, S., Remenyi, T. A., White, C. J. & Johnston, F. H. Heatwave and health impact research: A global review. *Health Place* **53**, 210–218 (2018).

18. Pyrina, M. & Domeisen, D. I. V. Subseasonal predictability of onset, duration, and intensity of European heat extremes. *Q J. R. Meteor Soc.* **149**, 84–101 (2023).

19. Lin, X., Wang, Y. J. & Song, L. C. Urbanization amplified compound hot extremes over the three major urban agglomerations in China. *Geophys. Res. Lett.* **51**, https://doi.org/10.1029/2023GL106644 (2024).

20. An, N. & Zuo, Z. Y. Changing structures of summertime heatwaves over China during 1961-2017. *Sci. China Earth Sci.* **64**, 1242–1253 (2021).

21. Zhao, L., Lee, X., Smith, R. B. & Oleson, K. Strong contributions of local background climate to urban heat islands. *Nature* **511**, 216–219 (2014).

22. Mohammad Harmay, N. S., Kim, D. & Choi, M. Urban Heat Island associated with Land Use/Land Cover and climate variations in Melbourne, Australia. *Sustain. Cities Soc.* **69**, 102861 (2021).

23. Chakraborty, T., Venter, Z. S., Qian, Y. & Lee, X. Lower urban humidity moderates outdoor heat stress. *AGU Adv.* **3**, e2022AV000729 (2022).

24. Huang, F. et al. Mapping local climate zones for cities: A large review. *Remote Sens Environ.* **292**, 113573 (2023).

25. Li, Y. F., Schubert, S., Kropp, J. P. & Rybski, D. On the influence of density and morphology on the Urban Heat Island intensity. *Nat. Commun.* **11**, 2647 (2020).

26. Shao, L. D. et al. Drivers of global surface urban heat islands: Surface property, climate background, and 2D/3D urban morphologies. *Build Environ.* **242**, 110581 (2023).

27. Yuan, Y. F., Liao, Z., Zhou, B. Q. & Zhai, P. M. Unprecedented Hot Extremes Observed in City Clusters in China during Summer 2022. *J. Meteorol. Res.-Prc.* **37**, 141–148 (2023).

28. Sun, Y., Zhang, X. B., Ren, G. Y., Zwiers, F. W. & Hu, T. Contribution of urbanization to warming in China. *Nat. Clim. Change* **6**, 706–709 (2016).

29. Rajeswari, J. R. et al. Urban heat island phenomenon in a desert, coastal city: The impact of urbanization. *Urban Clim.* **56**, 102016 (2024).

30. Wang, B., Liu, J., Kim, H.-J., Webster, P. J. & Yim, S.-Y. Recent change of the global monsoon precipitation (1979–2008). *Clim. Dyn.* **39**, 1123–1135 (2011).

31. Zhao, N. et al. Estimating the effect of urbanization on extreme climate events in the Beijing-Tianjin-Hebei region, China. *Sci. Total Environ.* **688**, 1005–1015 (2019).

32. Aram, F., García, E. H., Solgi, E. & Mansournia, S. Urban green space cooling effect in cities. *Heliyon* **5**, https://doi.org/10.1016/j.heliyon.2019.e01339 (2019).

33. Mohajerani, A., Bakaric, J. & Jeffrey-Bailey, T. The urban heat island effect, its causes, and mitigation, with reference to the thermal properties of asphalt concrete. *J. Environ. Manag.* **197**, 522–538 (2017).

34. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining,* 785–794 (2016).

35. Breiman, L. Random foresets. *Mach. Learn.* **45**, 5-32 (2001).

36. Tzavali, A., Paravantis, J. P., Mihalakakou, G., Fotiadi, A. & Stigka, E. Urban Heat Island Intensity: A Literature Review. *Fresen Environ. Bull.* **24**, 4535–4554 (2015).

37. Morabito, M. et al. Surface urban heat islands in Italian metropolitan cities: Tree cover and impervious surface influences. *Sci. Total Environ.* **751**, 142334 (2021).

38. Yang, X. Y., Li, Y. G., Luo, Z. W. & Chan, P. W. The urban cool island phenomenon in a high-rise high-density city and its mechanisms. *Int J. Climatol.* **37**, 890–904 (2017).

39. Li, H. D., Sodoudi, S., Liu, J. F. & Tao, W. Temporal variation of urban aerosol pollution island and its relationship with urban heat island. *Atmos. Res.* **241**, 104957 (2020).

40. Li, D. et al. Urban heat island: Aerodynamics or imperviousness?. *Sci. Adv.* **5**, eaau4299 (2019).

41. Ryu, Y. H. & Baik, J. J. Quantitative analysis of factors contributing to urban heat island intensity. *J. Appl Meteorol. Clim.* **51**, 842–854 (2012).

42. Steeneveld, G. J., Koopmans, S., Heusinkveld, B. G., van Hove, L. W. A. & Holtslag, A. A. M. Quantifying urban heat island effects and human comfort for cities of variable size and urban morphology in the Netherlands. *J. Geophys Res-Atmos.* **116**, D20 (2011).

43. Liu, J. P., Ren, Y. Q., Tao, H. & Shalamzari, M. J. Spatial and temporal variation characteristics of heatwaves in recent decades over China. *Remote Sens-Basel* **13**, 3824 (2021).

44. Wicker, W., Harnik, N., Pyrina, M. & Domeisen, D. I. V. Heatwave location changes in relation to rossby wave phase speed. *Geophys. Res. Lett.* **51**, e2024GL108159 (2024).

45. Luo, M. et al. Anthropogenic forcing has increased the risk of longer-traveling and slower-moving large contiguous heatwaves. *Sci. Adv.* **10**, eadl1598 (2024).

46. Luo, M. et al. Two different propagation patterns of spatiotemporally contiguous heatwaves in China. *Npj Clim. Atmos. Sci.* **5**, 89 (2022).

47. Orlanski, I. A. Rational subdivision of scales for atmospheric processes. *B Am. Meteorol. Soc.* **56**, 527–530 (1975).

48. Wang, K., Li, Y., Wang, Y. & Yang, X. On the asymmetry of the urban daily air temperature cycle. *J. Geophys. Res.: Atmos.* **122**, 5625–5635 (2017).

49. Yang, S. et al. Quantification of the urbanization impacts on solar dimming and brightening over China. *Environ. Res Lett.* **17**, 084001 (2022).

50. Li, M. M., Koks, E., Taubenböck, H. & van Vliet, J. Continental-scale mapping and analysis of 3D building structure. *Remote Sens Environ.* **245**, 111859 (2020).

51. Zhang, T. et al. A global dataset of daily maximum and minimum near-surface air temperature at 1 km resolution over land (2003–2020). *Earth Syst. Sci. Data* **14**, 5637–5649 (2022).

52. Guan, Y., et al. Elevation regulates the response of climate heterogeneity to climate change. *Geophys. Res. Lett.* **51**, e2024GL109483 (2024).

53. Yang, J., Xin, J., Zhang, Y., Xiao, X. & Xia, J. C. Contributions of sea–land breeze and local climate zones to daytime and nighttime heat island intensity. *npj Urban Sustainability* **2**, 12 (2022).

54. Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E. & Houston, T. G. An overview of the global historical climatology network-daily database. *J. Atmos. Ocean Tech.* **29**, 897–910 (2012).

55. Gong, P. et al. Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sens Environ.* **236**, 111510 (2020).

56. Hersbach, H. et al. The ERA5 global reanalysis. *Q J. R. Meteor Soc.* **146**, 1999–2049 (2020).

57. Liang, S. L. et al. The global land surface satellite (GLASS) product suite. *B Am. Meteorol. Soc.* **102**, E323–E337 (2021).

58. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777 (2017).

Acknowledgements

Author contributions

P.J., Conceptualization, Methodology, Formal analysis, Visualization, Writing-original draft preparation. X.Z., Data curation, Visualization. X.Y., Conceptualization, Funding, Investigation and Writing-review and editing. Q.X., Data curation, editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-025-02315-z.

Correspondence and requests for materials should be addressed to Xing Yuan.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.