PLOS Climate

RESEARCH ARTICLE

Evaluating the potential for heat warning systems to account for intra-urban variability

Svenja Ludwig[1]*, **Ferdinand Briegel**[2], **Andreas Christen**[1]

1 Chair of Environmental Meteorology, Faculty of Environment and Natural Resources, University of Freiburg, Freiburg im Breisgau, Germany, **2** IMKTRO, Regional Climate and Weather Hazards, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany

* svenja.ludwig@meteo.uni-freiburg.de

Abstract

Urban heat increasingly threatens public health as climate change intensifies heatwaves frequency, duration and severity. Germany's current county-level heat warning system might overlook urban–rural and intra-urban differences in heat exposure, limiting its usefulness for local action. This study developed and evaluated neighborhood-scale Pedestrian Heat Stress Products (PHSP) using high-resolution Universal Thermal Climate Index (UTCI) predictions from a machine learning framework. The Human Thermal Comfort Neural Network (HTC-NN) was used to generate hourly 1 × 1 m UTCI maps for Freiburg, Germany, during June–August 2023, accounting for buildings, vegetation, and street geometry. Predictions were aggregated to neighborhood scale (512 × 512 m), excluding buildings to represent pedestrian conditions. Daytime hours exceeding UTCI thresholds of 32 °C (strong heat stress) and 38 °C (very strong) were summed per 1 × 1 m cell and averaged by neighborhood. PHSP Level 1 was assigned when ≥2 hours exceeded 32 °C UTCI, Level 2 when ≥30 minutes exceeded 38 °C UTCI. PHSP demonstrated robust accuracy against official warnings (70.1% urban) and weather station data (82.0% urban). Comparing PHSP with official warnings revealed major differences: the current system missed 24 urban warning days in 2023. Urban areas experienced 1.2 times more Level 1 and 1.8 times more Level 2 days than rural surroundings, highlighting urban heat amplification and strong spatial heterogeneity linked to morphology. This study shows that PHSPs allow for differentiated warnings and can be used to transition from coarse, county-level to high-resolution, neighborhood-scale warning systems to deliver actionable, location-specific heat risk information.

OPEN ACCESS

Citation: Ludwig S, Briegel F, Christen A (2026) Evaluating the potential for heat warning systems to account for intra-urban variability. PLOS Clim 5(6): e0000941. https://doi.org/10.1371/journal.pclm.0000941

Editor: Phil McManus, University of Sydney, AUSTRALIA

Received: January 23, 2026

Accepted: May 8, 2026

Published: June 3, 2026

Data availability statement: The underlying PHSP data is available at https://doi.org/10.5281/zenodo.18301101. The HTC-NN code is available at https://doi.org/10.5281/zenodo.7974472. The official heat warning data are openly available at the Climate Data Center of the DWD.

Introduction

Urban heat has emerged as one of the most pressing public health challenges in Germany, with climate change intensifying the frequency, duration, and severity of extreme heat events globally [1].While the global annual mean air temperature (T_a)

Funding: SL received a stipend by the Hanns-Seidel-Stiftung. AC and SL acknowledge funding from the European Union's Horizon Europe program through the project "urbanAIR" (grant no. 101188131). FB received funding by the Helmholtz program "Changing Earth" at KIT. The ERC Synergy Grant 'urbisphere', funded by the European Research Council (ERC-SyG) within the European Union's Horizon 2020 research and innovation program under grant no. 855005, supported the collection of validation data (sensor network) and model tools used in this research (to AC and SL). The HTC-NN development was funded by the Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz (grant no. 67KI2029A/B, to AC and FB). We acknowledge support by the Open Access Publication Fund of the University of Freiburg. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

over land increased by 1.59 K over the last decade (2011–2020) compared to average annual T_a from 1850 to 1900, it increased by 2.0 K in Germany [2]. Cities are particularly vulnerable to heat stress due to their unique physical characteristics, including complex three-dimensional morphometry, extensive sealed surfaces, limited vegetation coverage, and significant anthropogenic heat emissions [3,4]. These factors fundamentally alter urban energy and radiation balances, creating substantial spatial and temporal variability in outdoor heat stress between urbanized areas and their rural surroundings as well as within urbanized areas [5,6]. Urban areas further experience increased surface roughness, which slows down regional winds while creating complex microscale airflow patterns, including sheltering, channeling along streets, and turbulent gusts near buildings [7,8]. Atmospheric moisture can also be modified by cities, with cities typically being drier during the day due to reduced evaporation from impervious surfaces but often becoming slightly more humid at night due to anthropogenic sources and reduced dew formation [4]. For a person outdoors, the built environment creates a highly variable radiation field; primarily due to the mix of sunlit and shaded facets, combined with different longwave radiation emitted from surfaces of different surface temperatures like walls and pavement. This results in a significantly altered mean radiant temperature (T_{mrt}), which expresses the short- and long-wave radiation absorbed at the outer surface of the body [9–11]. These altered climatic conditions directly influence the thermal sensation and stress experienced by individuals within the city and are highly variable within the city. The public health implications of these conditions are severe. Germany experiences significant heat-related mortality with approximately 48 000 heat related deaths in the last decade (2014–2023) [12,13]. Even single heat wave events can cause substantial mortality; the July 2023 heat wave alone contributed approximately 1100 deaths [13]. Heat exposure significantly increases mortality risk for specific diseases, with cardiovascular and respiratory conditions being particularly affected. Heat increases cardiovascular mortality risk by 15–24% and respiratory mortality risk by 34% [14,15].

To quantify this complex human thermal sensation and stress, and to manage the associated health risks, human thermal comfort indices are essential tools, as they move beyond relying on T_a alone [16].

These indices integrate multiple meteorological parameters, such as T_a, humidity (RH), wind speed (ws), and the three-dimensional radiative environment represented as T_{mrt}. Among the most advanced and widely adopted indices is the Universal Thermal Climate Index (UTCI), developed through international collaboration to provide a physiologically relevant, globally applicable measure of thermal stress across all climates and seasons [17,18]. UTCI employs a sophisticated multi-node model of human thermoregulation combined with a temperature-adaptive clothing model, simulating the physiological response of an active, "average" person in varying environments, and classifies thermal stress on a multi-category scale. Its input parameters include T_a, T_{mrt}, ws, and RH, making it highly sensitive and effective for heat assessment, as demonstrated in studies linking UTCI values to increased mortality during heatwaves across Europe [2,19,20]. Other prominent indices include the Perceived Temperature (PT), which is based on the human energy budget model

("Klima-Michel"), accounts for acclimatization, and defines thresholds for issuing public heat warnings [21,22]. The selection and application of a thermal comfort index generally depends on its intended purpose, ranging from epidemiological studies and public warnings to urban planning [16,23].

Heat warnings are crucial for mitigating the negative health impacts of intense heat episodes [21]. These systems aim to alert the public and authorities, enabling interventions and preventive measures [21,22].

The global heterogeneity of Heat-Health Warning Systems (HHWSs) illustrates that no single, universal methodology exists [23,24]. The design of an optimal HHWS involves complex trade-offs, primarily between geographical scale and the trigger variables chosen. Regarding scale, highly localized, city-specific warnings (e.g., in Italy) effectively capture urban heat island effects, offering high local relevance [23]. Conversely, broader county- or regional-level warnings are simpler to issue but risk masking critical urban-rural and intra-urban variability. This limitation reduces their effectiveness for targeted local action, as they may fail to detect localized heat stress in vulnerable populations [25]. Regarding variables, similar trade-offs exist. Complex indices, like Germany's PT, offer greater physiological relevance by integrating humidity, wind, and radiation. Other systems, like the HHWSs in France and England, are strengthened by incorporating minimum T_a to account for the health impact of insufficient overnight cooling [23]. Additionally, the threshold determination depends on local climate as well as on the target application. Several methods for threshold determination exist such as percentile-based methods, categorical values of thermal indices or linking temperatures to mortality data [26–29]. This diversity, mirrored in North American systems [24], underscores that the approaches are highly dependent on local climatology, data availability, public health infrastructure, and the specific health impacts being targeted [23,24]. The reliance on coarse spatial scales and varied thresholds has measurable consequences.

Recent research across twelve European countries, including Germany, shows that official heat warning systems often underestimate the societal impact of heat waves [30]. Multiple independent indicators, ranging from online search trends to press coverage and excess mortality, reveal that populations respond to hot weather on far more days than those covered by formal heat warnings. In Germany, although the alignment between societal responses and official warnings is better than in other countries, substantial gaps remain, with many heat-affected days going unflagged. This underscores the need to look at localized urban heat effects, recalibrate thresholds and criteria so that warnings more closely reflect actual societal vulnerability and behavioral responses, particularly in urban contexts that are characterized by a high heterogeneity of outdoor heat stress.

The advancements in urban climate modeling and machine learning applications have opened new opportunities for developing high-resolution and computationally rapid heat stress assessment tools. In particular, studies such as [31–34] have explored the use of machine learning to predict outdoor thermal comfort at high spatial and temporal resolution across entire cities and urban-rural landscapes. Machine learning approaches, especially artificial neural networks trained on physically-based urban climate model outputs or on comprehensive measurement data, can rapidly generate fine-scale thermal comfort predictions across complex urban landscapes. Crucially, they are capable of capturing the nonlinear relationships between meteorological conditions, urban morphology, and thermal stress, while maintaining computational efficiency [31]. For outdoor human thermal comfort modelling, it is therefore important to know the scales of these urban effects. Where radiation and wind are extremely heterogenous within the city and need a high-scale representation (1 x 1 m), T_a and RH vary at different scales and can be modelled on a coarser, neighborhood scale, which saves resources [31]. However, despite these technological advances, the operational use of high-resolution urban climate and machine learning models in current heat warning systems is limited. For example, existing systems such as Germany's HHWS operate at coarse spatial resolutions, typically at county or regional scales.

Whereas current HHWS struggle to resolve differences in neighborhood-level risks, emerging research points toward the benefits of finer spatial detail. For instance, a personalized early warning system for Delhi has been developed that employs UTCI-based approaches to deliver user-specific hazard forecasts at 333 m spatial resolution [35]. These reveal critical spatial heterogeneities and local heat hotspots that would be obscured at coarser scales. Similarly, previous

research highlights that incorporating urban-scale realism is a core requirement for the next generation of HHWSs [25]. Nonetheless, critical questions remain regarding the magnitude and significance of intra-urban heat variability, the effectiveness of current county-scale systems in addressing it, and the optimal metrics and thresholds for fully spatially explicit warnings.

Research hypothesis

We hypothesize that that there is significant variability in outdoor human heat stress across urban and rural areas that is not adequately captured by the current county-scale HHWS in Germany. We also hypothesize that incorporating urban effects through high-resolution, physiologically-based thermal comfort metrics (UTCI) into spatially-explicit HHWSs will provide a more accurate and actionable assessment of heat risk for urban populations. Specifically, we expect that:

1. Neighborhood-scale heat stress will exhibit substantial spatial heterogeneity within cities and between cities and their surroundings.
2. Urban areas will demonstrate more frequent and intense outdoor heat stress conditions compared to surrounding rural areas, particularly during heat warning periods.
3. Current county-scale heat warnings will show varying accuracy across different urban environments and microclimates, with systematic under-warnings in urban neighborhoods.

This study addresses the critical gap between current county-scale heat warning systems and the need for neighborhood-scale outdoor heat stress warnings in urban environments. Our primary objectives are to:

- Develop physiologically-based heat stress criteria suitable for urban neighborhood-scale applications using UTCI thresholds linked to human thermal responses.
- Implement a prototype high-resolution heat stress prediction system using machine learning techniques to generate spatially-explicit thermal comfort assessments.
- Quantify urban-rural heat stress disparities by comparing neighborhood-scale heat stress patterns against official county-scale heat warnings.
- Assess spatial variability in warning accuracy across the study area to identify systematic biases in current warning systems.
- Evaluate the potential for improved heat risk assessment through spatially-explicit warning systems and determine optimal metrics for urban heat warning applications.

Materials and methods

A framework for neighborhood-scale daily Pedestrian Heat Stress Products (PHSP) was developed in this study. It encompasses three steps (Fig 1): (1) The application of the Human Thermal Comfort Neural Network (HTC-NN) forced with operational forecast data for obtaining highly spatial and temporal resolved UTCI maps, (2) an aggregation to daily neighborhood scale (512 x 512m) UTCI Exceedance Maps (e.g., hours per day with UTCI ≥ 32°C) and (3) in a last step the calculation of the PHSP at two warning levels and in daily resolution. The components are described in detail in the following sections.

Development of a prototype neighborhood-scale heat warning framework

Human Thermal Comfort Neural Network (HTC-NN). The foundation of our neighborhood-scale heat stress assessment system is the Human Thermal Comfort Neural Network (HTC-NN), a machine learning framework developed

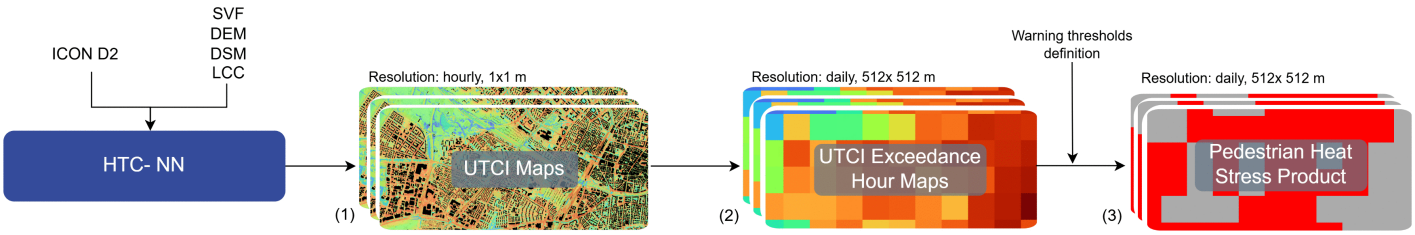

Fig 1. Schematic overview of the PHSP workflow.

https://doi.org/10.1371/journal.pclm.0000941.g001

by Briegel et al. [31] specifically for high-resolution urban thermal comfort prediction. Since its development the HTC-NN has been applied in thermal comfort mapping [31], long-term city-wide climate projections of heat stress [36], evaluation of satellite land surface temperature as a proxy for outdoor heat stress [37] and AI-supported urban adaptation planning [38]. However, its use within a heat warning framework is a novel application. The HTC-NN represents a significant advancement in computational thermal comfort assessment, combining the physical rigor of numerical urban climate modeling with the computational efficiency required for operational applications or downscaling climate ensembles to a building resolved scale [36]. The HTC-NN relies on two main categories of inputs: meteorological forcing data and high-resolution geospatial data.

Meteorological inputs include T_a, RH, ws, wind direction (wd), atmospheric pressure, and downwelling radiation fluxes. The meteorological forcing data was obtained from historical weather forecasts of the regional-scale numerical weather prediction (NWP) model ICON D2 operated by the German Weather Service (DWD) [39]. Data for the next 24 h were used from the ICON D2 run at 00.00 UTC. In 2023, ICON D2 was run without the urban parametrization Terra-URB that went operational at DWD only in February of 2025 [40]. The forecast data has an hourly resolution, and the study period covers the summer of 2023 (June-August).

Geospatial data describe the city's physical environment and include land cover classification, digital elevation and surface models (for both buildings and vegetation) [11,31]. By integrating four sub models for T_a, RH, T_{mrt}, and ws [41] respectively, the high-resolution, spatially explicit outputs were then used to calculate UTCI at a 1 x 1 m resolution [31].

Pedestrian Heat Stress Product (PHSP) framework for warnings. To provide sufficient spatial detail while capturing intra-urban variability and remaining conceptually accessible to users, the hourly predictions by the HTC-NN were aggregated from 1 x 1 m to neighborhood scale (512 x 512 m). As the focus is on pedestrian-level conditions, grid cells with buildings were excluded from the aggregating process. The number of hours exceeding UTCI 32°C h_{32} (strong heat stress) and hours exceeding UTCI 38°C h_{38} (very strong heat stress) within each 1 x 1 m grid cell during the daytime (after sunrise and before sunset) were summed up on a daily basis. Next, for each 512 x 512 m grid cell, h_{32} and h_{38} values were averaged from the 1 x 1 m data. The aggregated daily h_{32} and h_{38} maps then underwent the criteria checks for heat stress warnings. These PHSP warning criteria are based on description of UTCI threshold exceedances linked to documented physiological responses to heat stress [18,20]. A PHSP warning of level 1 is issued for a specific grid cell when its h_{32} value exceeds two hours. A UTCI value of at least 32 °C indicates situations of strong heat stress, which are associated with an increase in rectal temperature after 120 minutes, an instantaneous rise in skin temperature, and an average sweat rate above 200 g h^{-1}. Clinically, without countermeasures undertaken this magnitude of sustained fluid loss and core heating translates to hypovolemia, elevated risk of cardiovascular drift, heat syncope, acute kidney injury, and the progressive onset of heat exhaustion (see Table 1) [42,43]. A PHSP warning of level 2 is issued when h_{38} exceeds 30 minutes. Under these conditions, heat stress is more severe, characterized by a reduced core-to-skin temperature gradient and a rapid increase in rectal temperature within 30 minutes. This translates to profound cardiovascular strain, a high risk of acute cardiovascular collapse, and an imminent progression toward heat stroke (see Table 1) [44–46].

Table 1. PHSP level and criteria description, extracted from [18]. The health implications listed in this table are not directly observed outcomes of this study. They are linked consequences based on established clinical literature derived from the physiological responses associated with the UTCI thresholds and the corresponding exposure durations [42–46].

PHSP Level	UTCI Threshold	Criteria in grid cell	Physiological response	Health implications
L1	≥ 32°C	h_{32} >120 min	Average human sweat rate >200 g h^{-1}. Increase in rectal temperature at 120 min. Instantaneous change in skin temperature.	Sustained hypovolemia and dehydration. High risk of heat syncope (fainting), heat cramps, acute kidney injury and heat exhaustion.
L2	≥ 38°C	h_{38} >30 min	Low core-skin temperature gradient. Increase in rectal temperature at 30 min.	Severe cardiovascular strain due to massive cutaneous vasodilation and compromised convective cooling. High risk of acute cardiovascular collapse and imminent progression to heat stroke.

https://doi.org/10.1371/journal.pclm.0000941.t001

The final PHSP map assigns each 512 x 512 m grid cell to one of three categories: L0 (no warning), L1 (strong heat stress), L2 (very strong heat stress) (see Fig 3).

Study area

The study was conducted in Freiburg im Breisgau, Germany, and surrounding areas, covering approximately 150 km² with elevations ranging from 200 to 400 m above sea level (Fig 2). Freiburg has approximately 230 000 inhabitants [49] and experiences a temperate oceanic climate with warm summers and regularly elevated heat stress conditions with over 20 hot days ($T_{a,max}$ > 30°C) per year in the last decade [50,51]. This climatic setting enables a robust and realistic analysis of heat stress conditions. Based on the Local Climate Zone (LCZ) classification, most of the urban study area falls into LCZ 5 (open mid-rise), LCZ 6 (open low-rise), and LCZ 8 (large low-rise), while a small part of the city centre is classified as LCZ 2 (compact mid-rise). For the rural parts of the study area the two main classes are LCZ A (dense trees) and LCZ C (bush, scrub) [52,53]. The urban landscape includes typical Central European urban forms: a city center with narrow streets and mid-rise buildings, suburban residential areas with varying building densities, industrial zones with large flat roofs, impermeable surfaces and minimal vegetation, and extensive forested and agricultural areas that provide important reference conditions for rural heat stress patterns. A key reason for selecting Freiburg as study area is the availability of a unique street-level biometeorological sensor network of 11 stations measuring including T_a, RH, ws, wd, and globe temperature (T_g) required to calculate the Universal Thermal Climate Index (UTCI) [51,54]. To our knowledge, such a network with this level of detail was not available elsewhere in 2023. This unique data availability enabled the training, evaluation, and validation of the HTC-NN model, resulting in a high level of confidence in the developed framework.

Study period

The study was conducted during the summer months of June, July and August (JJA) in 2023. This period was chosen due to coherent data availability (heat warning data, street-level observations, and historical weather forecast data) but also reflects the increasingly warmer climatic conditions currently experienced in Freiburg and expected under future climate change. 2023 was recorded to be the warmest year to date in Freiburg with an annual mean air temperature of 12.9 °C (+1.4°C above the 1991–2020 climatic reference period), and the fifth warmest JJA period (+1.1°C) [55].

Evaluation of the PHSP

Comparison of PHSP warnings to warnings based on observational data. To further evaluate the PHSP warning approach, we compared PHSP warnings against warnings based on observational data (using the same criteria as in PHSP) from Freiburg's street-level weather station network described by [48,51,54]. Eleven measurement stations record

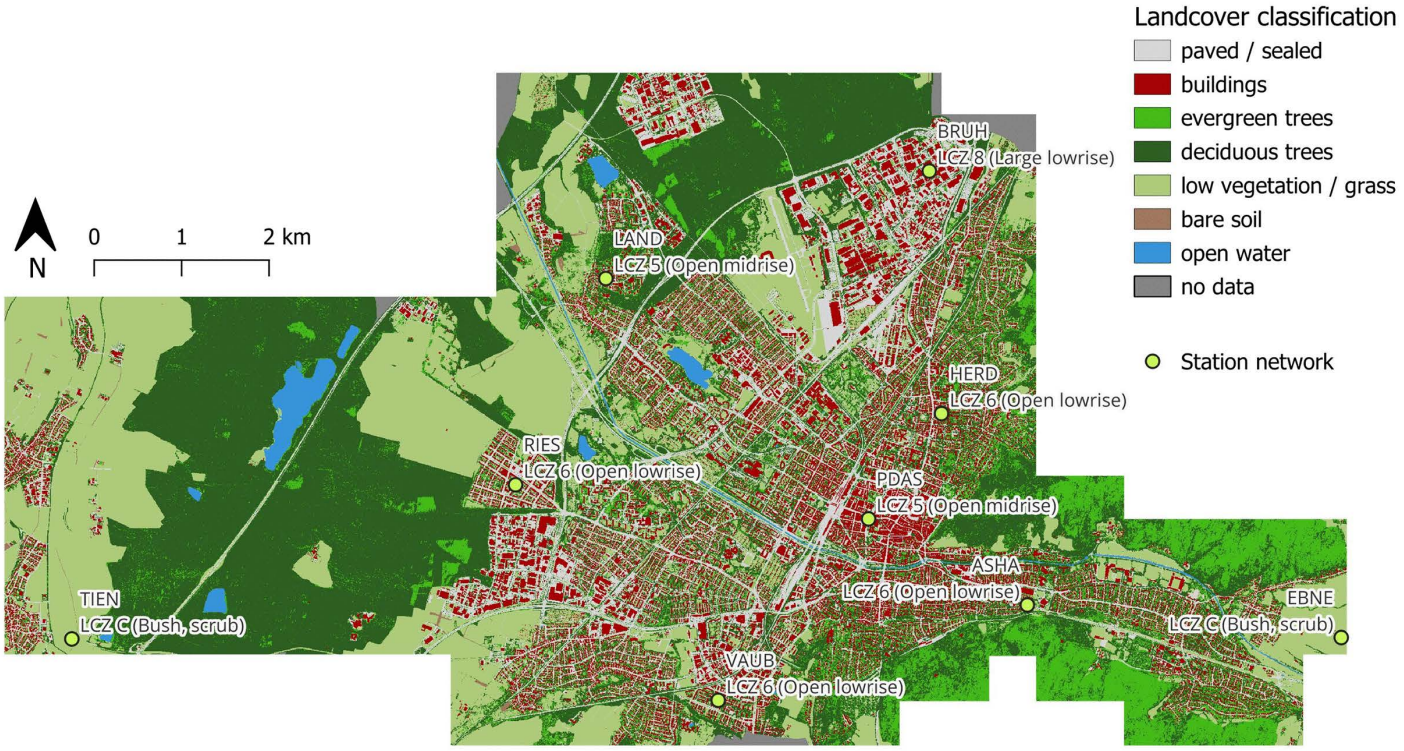

Fig 2. Map of study area with landcover classification [47] and weather station network [48]. Even though rural station EBNE is not located within the study domain, is considered to be representative for the next closest PHSP cell (distance approx. 230 m) due to same LCZ and landcover (open field).

https://doi.org/10.1371/journal.pclm.0000941.g002

all parameters required for determining UTCI, including T_a, RH, ws, wd, and T_g every 30 seconds. The two instruments (ClimaVUE50 and BLACKGLOBE-L, both by Campbell Scientific Inc) are installed at 3m height. The detailed exposure information per station can be found in the network documentation [48]. Of the 11 stations, 9 stations are within the study area and measured data for the study period. The 9 stations of the network are strategically distributed across urban and rural environments: 7 urban stations located in various urban morphological contexts and 2 rural stations in representative rural settings (see Table 2). The observational data underwent quality checks in accordance with plausible value ranges: T_a: -35–45 °C, downwelling shortwave radiation (rsds): 0–1300 W/m, T_g: -20–60 °C, RH: 10–100%, ws: 0–40 m/s. Faulty values were removed and handled as data gaps; no gap filling was performed [56].

T_{mrt} values were calculated using Equation 1 [57] and the manufacturer information of the black globe [58].

$$T_{mrt} = \left[\left(T_g + 273.15\right)^4 + \frac{1.1 \times 10^8 \times ws^{0.6}}{\varepsilon \times D^{0.4}} \times \left(T_g - T_a\right) \right]^{\frac{1}{4}} - 273.15 \quad (1)$$

With T_g = Globe temperature (°C), ws = Wind speed, T_a = Air temperature (°C), D = Diameter of the black sphere (m) = 0.152 m and ε = Emissivity of the black sphere = 0.957.

Hourly UTCI values were then calculated on the basis of T_{mrt} following [31]. The code is available at Briegel (2023) [59].

Then we applied the same criteria from the PHSP framework to the observation data (see Section 2.1). The resulting observational-based warnings were compared against PHSP warnings for corresponding grid cells enabling assessment of both PHSP framework accuracy and spatial representativeness. To evaluate the performance of the PHSP warnings

Table 2. List of weather stations within the study area. Further description of stations can be found in Plein et al. [48], * = rural station.

Station	LCZ	Dominant land use	Sky View Factor	Elevation (amsl)
ASHA	6 (Open lowrise)	Residential	0.72	296.0 m
BRUH	8 (Large lowrise)	Industrial	0.73	237.6 m
EBNE*	C (Bush, scrub)	Agricultural	0.90	340.3 m
HERD	6 (Open lowrise)	Residential	0.52	265.8 m
LAND	6 (Open lowrise)	Residential	0.70	234.2 m
PDAS	5 (Open midrise)	Recreational/Commercial	0.68	279.3 m
RIES	6 (Open lowrise)	Residential	0.66	237.3 m
TIEN*	C (Bush, scrub)	Agricultural	0.84	210.8 m
VAUB	6 (Open lowrise)	Residential	0.58	258.7 m

https://doi.org/10.1371/journal.pclm.0000941.t002

for the stations and their corresponding 512 x 512 m grid cells the metrics accuracy, per-class F1 and macro F1 [60] were calculated via the following equations:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$False\ Alarm\ Rate\ (FAR) = \frac{FP}{FP+TN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision+Recall} \tag{6}$$

$$Macro\ F1 = \frac{1}{n}\sum_i F1_i = \frac{1}{n}\sum_i \frac{2 \times Precision \times Recall}{Precision+Recall} \tag{7}$$

with TP = number of true positives, TN = number of true negatives, FP = number of false positives, FN = number of false negatives and n = number of classes. The accuracy gives the proportion of correct classified warning days to the overall number of days. The False Alarm Rate (FAR) is the ratio between false positives and the total number of actual negative events. Precision measures the correctness of the predictions for a class, while Recall measures the completeness. F1 score is the harmonic mean of Precision and Recall. The macro F1 score allows to evaluate the average performance over all classes of a multi-class classification model, while treating them equally [60].

Comparison to official heat warning products. The official German HHWS operates at county scale, using numerical weather forecasts based on the ICON D2 model [23], to generate automated heat warnings that must be confirmed by trained biometeorology forecasters before public release [21]. This human oversight allows for adjustments based on local experience and additional context before warnings are disseminated. The system calculates PT using a standardized reference person called "Klima-Michel," modeled as a 35-year-old male walking at 4 km/h [61]. The PT

calculation incorporates T_a, ws at 1.1 m, RH, T_{mrt}, and personal-physiological factors such as clothing and physical activity [21]. Heat warnings are issued for the current and following day when PT values at 12:00 UTC reach strong heat stress levels of PT ≥ 32 °C or higher (L1), or extreme heat stress levels of PT ≥ 38 °C or higher (L2). The system features dynamic threshold adjustments that account for short-term acclimatization: Warning thresholds automatically adjust based on PT values from the previous 30 days, with thresholds increasing up to a maximum of 34 °C PT for strong heat load when sustained heat has occurred. This prevents excessive warning issuance during extended heat episodes. Additionally, thresholds vary geographically to reflect regional climate differences, with lower values applied in northern Germany's cooler climate and higher values in the warmer southern regions [21].

Special considerations for vulnerable populations include modeling elderly vulnerability through a modified reference person called "Klima-Michel Senior", represented as a 75-year-old individual with reduced activity and body mass. When this model indicates extreme heat stress before general population thresholds are met, warnings include specific advices for older adults. The system also incorporates a building energy model framework to estimate nighttime operative room temperatures in a standard building, considering different room orientations and sleep quality concerns in warning decisions [62]. For cities with populations exceeding 100 000, the model accounts for the urban heat island effect, which enhances nighttime temperatures. Weather predictions are adjusted to incorporate this effect before calculating PT and simulated indoor temperatures [21].

Historical official heat warning datasets are openly available for Germany [63]. The study area falls within only one warning region; the county of Breisgau-Hochschwarzwald including the city of Freiburg with an area of 1531.3 km^2 [64]. During summer of 2023, DWD issued heat warning levels L1 and L2, 10 and 11 times, respectively for the county of Breisgau-Hochschwarzwald. To assess the validity and practical applicability of the PHSP framework, we compared the PHSP warnings against the official DWD heat warning system for the JJA period of 2023.

To bridge the difference in spatial resolution, where the DWD provides a singular daily value for the entire study extent, the PHP framework provides grid-based outputs, a pixel-wise comparison was conducted. The official warning status of each day was cross-referenced against the output of every individual 512 x 512 m PHSP grid cell. For this analysis, cells were categorized based on building density: urban areas were defined as grid cells with a building fraction of at least 15%, while rural areas were those below this threshold. Performance metrics were derived by calculating the spatial average of matching warning events across all cells within these respective classes. Comprehensive statistical analyses were conducted to quantify the performance and characteristics of the PHSP framework in urban and rural areas, respectively. Confusion matrices were generated and the metrics similarity and warning bias (see Equations 8 and 9) were calculated.

$$Similarity = \frac{1}{T}\sum_{d=1}^{T} I(L_{PHSP}(d) = L_{DWD}(d)) \tag{8}$$

$$Warning\ bias = \frac{1}{T}\sum_{d=1}^{T} I(L_{PHSP}(d) - L_{DWD}(d) \geq 1 \tag{9}$$

with d = a specific day, from 1 to T = total days, $L_{PHSP}(d)$ = the classification value from PHSP on day d (0, 1, or 2), $L_{DWD}(d)$ = the classification value from DWD on day d (0, 1, or 2), I = the indicator function (1 if true, 0 if false). Similarity ranges from 0 (no agreement) to 1 (perfect agreement). Warning bias ranges from 0 (perfect agreement) to 1 (PHSP always detects at least 1 level higher than DWD), thus indicating a positive bias.

Additionally, spatial analysis to characterize the geographic distribution of similarity and systematic warming bias was performed. We calculated and mapped the percentage of days when PHSP warnings matched official DWD warnings for each grid cell, enabling identification of areas where county-scale warnings are most and least representative of local conditions.

Results

Spatial overview of PHSP warning days

Heat stress in the summer of 2023 was ubiquitous across the entire study area but significantly intensified and more uniform within urban grid cells compared to rural landscapes.

This trend is evident in the spatial distribution and frequency of detected warning days. For the summer period (JJA) of 2023, 92 maps with grid cell size 512 x 512 m were calculated. An example with a PHSP warning for a selected day, June 25, 2023, can be seen in Fig 3. On this day, almost all grid cells within the study area were classified at least PHSP L1. The cells were PHSP L2 was detected are mostly located within the hatched urban area.

The total number of detected PHSP heat warning days per grid cell in the summer of 2023 is shown in Fig 4. The minimum number of PHSP L1 warning days was 14, compared to 21 official DWD warning days within the county. However, the majority of grid cells (94%) experienced more than 21 PHSP L1 warning days. Urban grid cells showed up to a maximum of 55 days with PHSP L1 warning days. The distribution of PHSP L2 warnings was similar, with a maximum of 20 days in urban grid cells. Only 7 grid cells did not experience any PHSP L2 warning, all located in forests.

The split violin plots shown in Fig 5 highlight an intensification of heat stress in built-up areas. Urban grid cells consistently experienced more heat stress days than rural cells for both levels. For urban grid cells a mean day count of 27.4 for L1 and 17.4 for L2 can been observed. Rural grid cells record a lower average day count for L1 of 22.8 and for L2 of 9.6. The shape of the distributions indicates differences in variability: the rural class displays a wide, dispersed range (Standard deviations: 4.2 (L1) and 3.9 (L2)) reflecting a heterogeneous pattern. Conversely, the urban class shows a more narrow distribution with lower standard deviation (3 for L1, 1.4 for L2), indicating that heat stress levels are consistently high and more uniform across the urban grid cells.

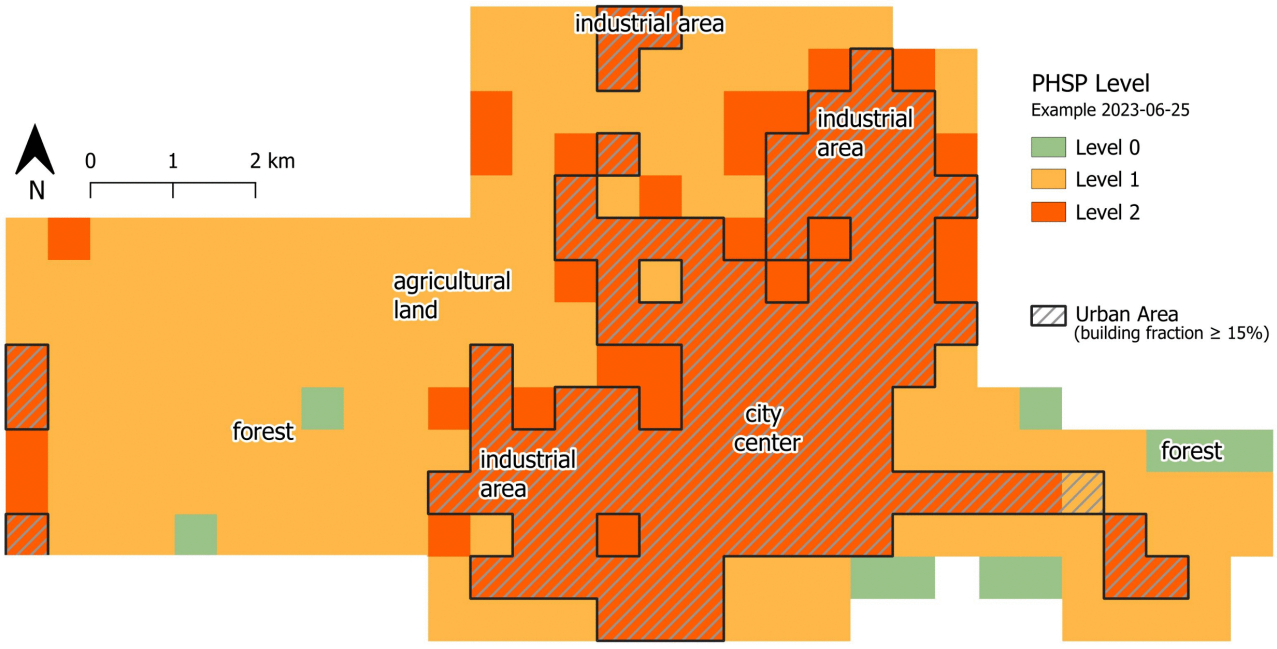

Fig 3. Map of study area showing the different PHSP heat warnings for a selected day (2023-06-25).

https://doi.org/10.1371/journal.pclm.0000941.g003

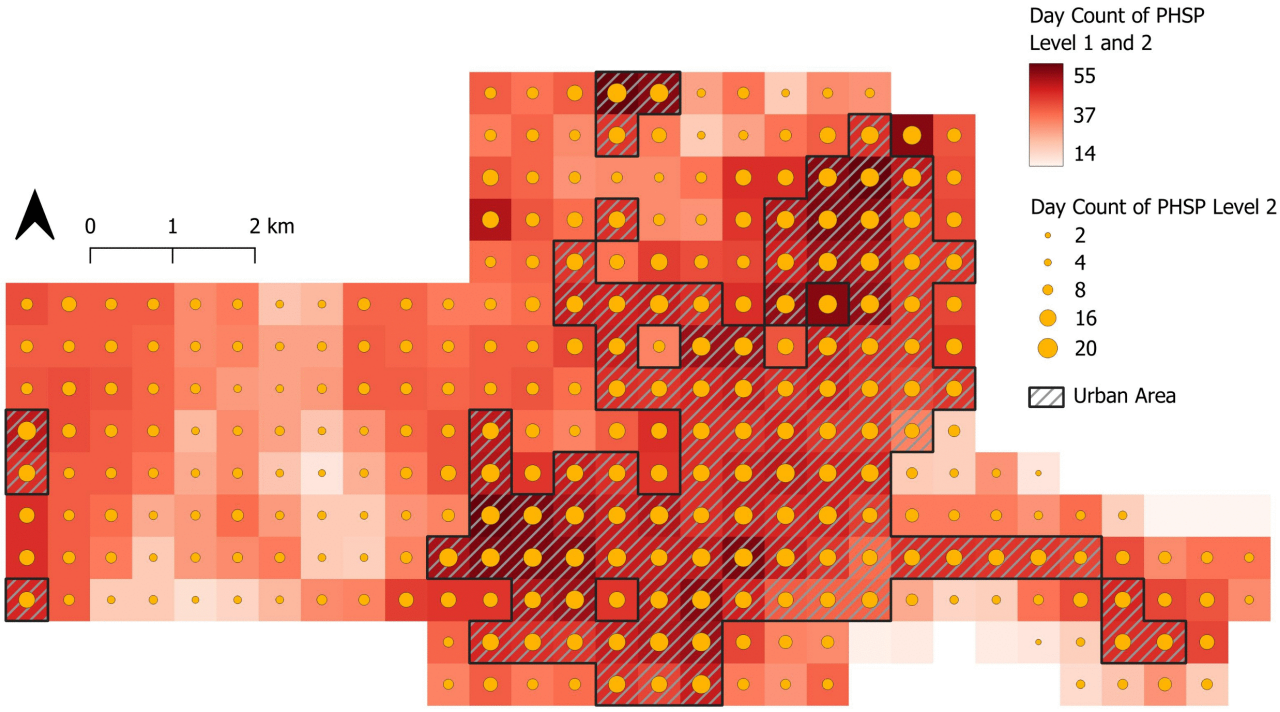

Fig 4. Map of study area showing the count of PHSP heat warning days per 512 x 512 m grid cell in the period JJA 2023 (total: 92 days). The overall PHSP warning day count is shown on the color scale, while the count of PHSP Level 2 warning days is shown by circle size.

https://doi.org/10.1371/journal.pclm.0000941.g004

Comparison of PHSP warnings to warnings based on observational data

This section reports how these PHSP warnings compared against warnings derived from the street-level sensor network. While urban stations generally showed a higher frequency of heat warning days, the PHSP matched these days with high accuracy, whereas its predictive performance at rural stations was lower.

To establish this, the PHSP warnings were compared to warnings based on the 9 street-level weather stations and the PHSP framework. For both warning levels, urban locations recorded more heat days than rural locations. For L1, urban stations averaged 40 days (corresponding PHSP: 28) compared to rural stations with 29 days (corresponding PHSP: 26.5). This trend continued for L2, where urban stations averaged 16 days (PHSP: 18) versus rural stations with 9 days (PHSP: 8).

The comparison of the PHSP warnings against warnings based on observational station data revealed a notable disparity in performance between urban and rural environments (see Table 3). The PHSP warnings were robust and have a consistent accuracy for urban stations, achieving an average accuracy of 82.0% and a macro F1 score of 0.82. The confusion matrices for urban sites, such as ASHA and BRUH, showed strong agreement along the diagonal, indicating a high rate of correct classifications across all three levels (L0, L1, and L2, see Fig 6). In contrast, the PHSP warning performance in rural areas was considerably lower and more variable, with an average accuracy of 75.6% and a macro F1 score of just 0.7. This underperformance was exemplified by the rural station TIEN, where an accuracy of only 68.5% was recorded, and where significant misclassification was observed, particularly for L0 (F1 score of 0.44).

Comparison of PHSP warnings with official DWD warnings

In addition to this station-based comparison, the PHSP framework was compared to the official DWD heat warning framework for the county of Breisgau-Hochschwarzwald, which entails the entire study area. In contrast to the high agreement

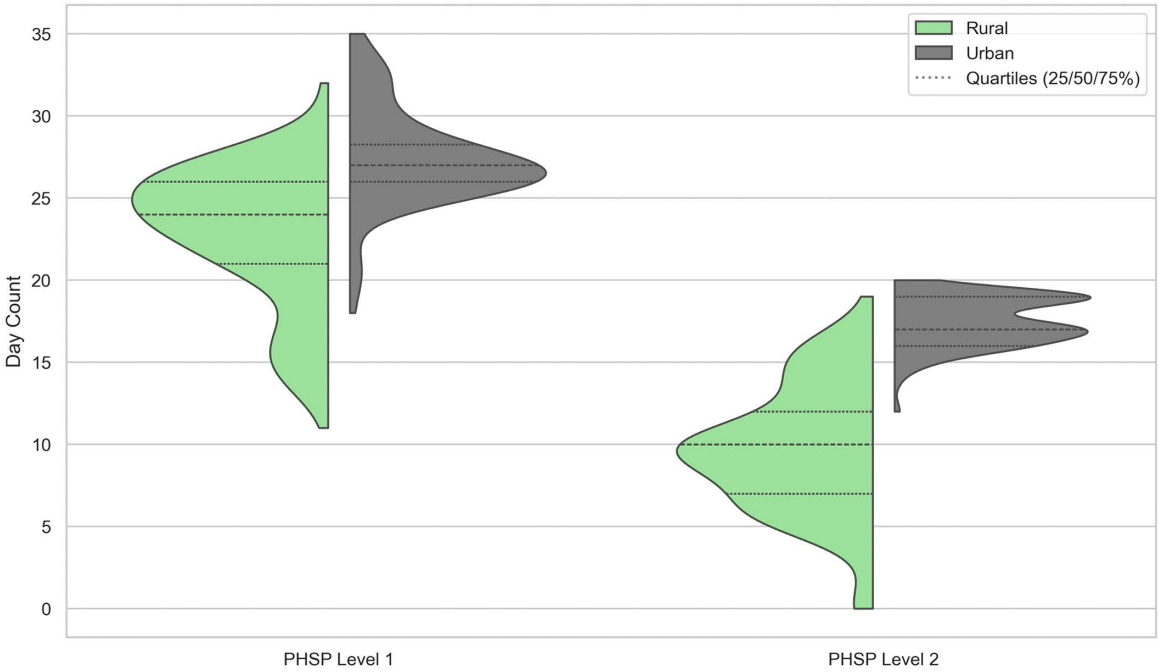

Fig 5. Distribution of day count for PHSP Level 1 (left) and Level 2 (right) for all grid cells in the study area and the entire study period. Data are classified in rural (green) and urban (grey) grid cells. The dotted lines show the interquartile range (IQR, 25–75%).

https://doi.org/10.1371/journal.pclm.0000941.g005

Table 3. Precision metrics of PHSP warnings determined by comparison with warnings based on observational data of the sensor network, distinguished between urban and rural areas.

Type	Station	Matches	Mismatches	Accuracy (%)	F1			Macro F1
					L0	L1	L2	
urban	ASHA	72	20	78.3	0.81	0.71	0.88	0.80
	BRUH	81	11	88.0	0.89	0.85	0.92	0.89
	HERD	74	18	80.4	0.85	0.73	0.83	0.80
	LAND	75	17	81.5	0.85	0.74	0.88	0.82
	PDAS	76	16	82.6	0.87	0.77	0.84	0.82
	RIES	77	15	83.7	0.85	0.78	0.87	0.84
	VAUB	73	19	79.3	0.80	0.75	0.91	0.80
	Average	**75.4**	**16.6**	**82.0**				**0.82**
rural	EBNE	76	16	82.6	0.90	0.72	0.67	0.76
	TIEN	63	29	68.5	0.77	0.53	0.59	0.63
	Average	**69.5**	**22.5**	**75.6**				**0.70**

https://doi.org/10.1371/journal.pclm.0000941.t003

with local sensors, the PHSP framework significantly diverges from official county-wide DWD warnings by identifying localized urban heat events that the regional system misses.

On average, rural areas demonstrated substantially better agreement with official DWD warnings, achieving an overall similarity of 78.8% compared to 70.1% in urban areas (Fig 7). The macro F1 score was 0.64 for urban areas and 0.67 for rural areas (Table 4), indicating moderate classification agreement. The averaged PHSP warnings in rural areas better matched the DWD warnings, mainly by being excellent in identifying days without any heat warnings (L0, F1 score = 0.89).

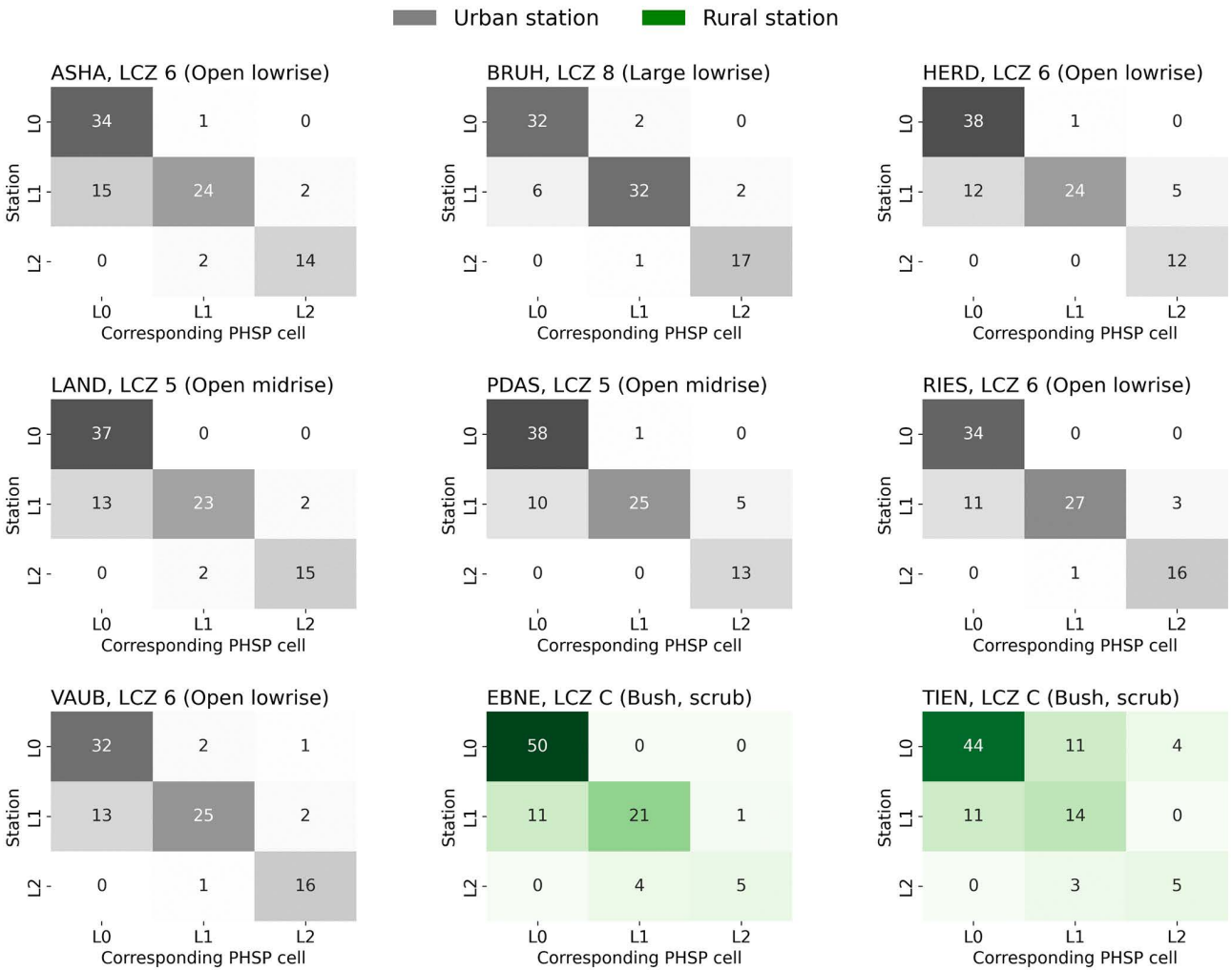

Fig 6. Confusion matrices comparing PHSP levels (L0, L1 and L2) for each grid cell against their corresponding warnings based on observational data from the sensor network for the period JJA 2023.

https://doi.org/10.1371/journal.pclm.0000941.g006

Overall, the PHSP framework issued on average 24 warning days for urban areas (20.8 L1, 3.2 L2) even though the DWD did not issue any official warnings.

In both rural and urban contexts, there was low similarity between the PHSP framework and the official DWD warnings for L1 warnings. In urban areas, when PHSP suggested a L1 warning, it only aligned with an official DWD warning 24% of the time (precision), and its high FAR (25.5%) indicates it frequently issued a L1 warning when the DWD did not.

When looking at L2 warnings, the PHSP showed different detection precision for urban and rural contexts entirely. In urban areas, PHSP correctly identified nearly all DWD L2 warnings (0.99 recall), ensuring that no critical event went unnoticed. Yet it also issued L2 warnings on some days when the DWD did not (FAR of 8.1%). In contrast to urban areas, the PHSP warnings showed a high precision (0.76) to DWD L2 warnings in rural areas, meaning that the detected L2 days are highly reliable. However, a recall of 0.67 means that more some of the official DWD warnings couldn't be detected in rural settings. In summary, the PHSP system issues more L2 warnings for urban areas and fewer for rural areas than the official warnings issued by the DWD. Additionally, it is to be mentioned, that the first warning day also differed between

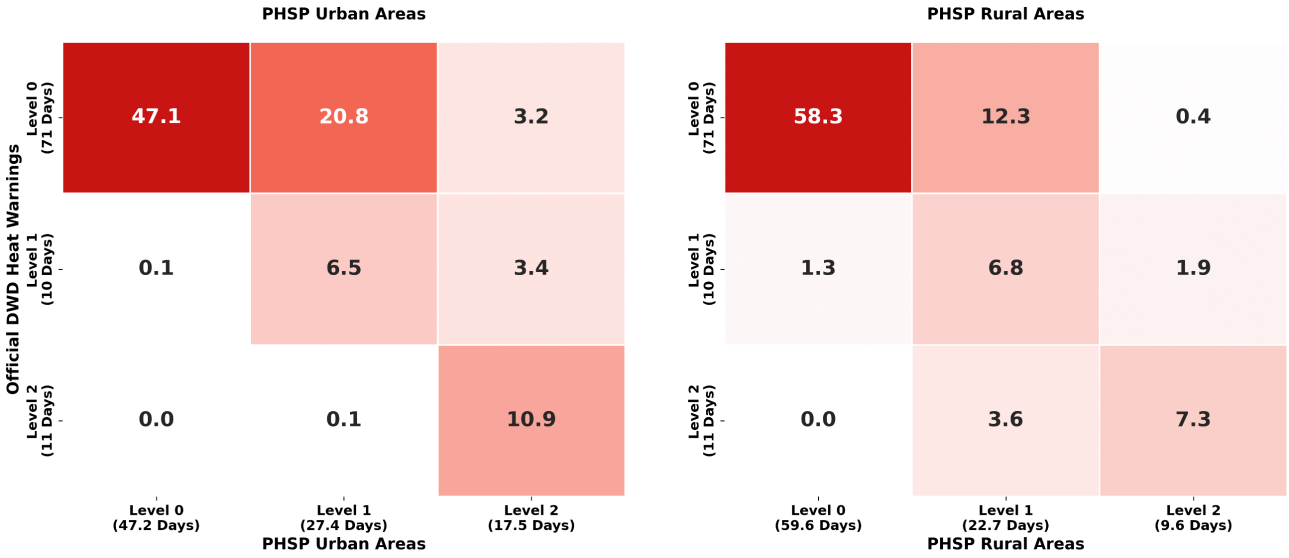

Fig 7. Confusion matrices comparing PHSP warnings against DWD warnings from JJA 2023 for averaged urban (left) and averaged rural (right) contexts.

https://doi.org/10.1371/journal.pclm.0000941.g007

Table 4. Precision metrics of PHSP framework determined by comparison with the official DWD warning system distinguished between urban and rural areas.

	Class	FAR (%)	Precision	Recall	F1 Score	Macro F1	Similarity (%)
Urban	L0	0.5	1.00	0.66	0.80	0.64	70.1
	L1	25.5	0.24	0.65	0.35		
	L2	8.1	0.62	0.99	0.77		
Rural	L0	6.2	0.98	0.82	0.89	0.67	78.8
	L1	19.4	0.30	0.68	0.42		
	L2	2.8	0.76	0.67	0.71		

https://doi.org/10.1371/journal.pclm.0000941.t004

PHSP and DWD as well as between PHSP in rural and in urban contexts. The first PHSP L1 warning for the general urban area (L1 for 90% of the urban grid cells) was detected on 2023-06-01, which is 18 days prior to the first official DWD L1 warning (2023-06-19). The first PHSP L2 warning for the general urban area (L2 for 90% of the urban grid cells) was on 2023-06-20, which is 19 days prior to the official DWD L2 warning (2023-07-09). The PHSP in the general rural areas indicated L1 warning only two days prior to official DWD L1 warning and first L2 warning on the same day (2023-07-09).

Intra-urban differences. Investigating the similarity of the PHSP framework and the county-wise official DWD warnings per grid cell, a distinct geographic pattern across the study area became visible (see Fig 8). Within the urban area, warning agreement with DWD was not uniform but varied systematically with urban morphology: areas of higher building fraction showed both the lowest similarity to official warnings and the highest positive warning bias, with industrial zones and the city center most affected.

On average, the similarity between PHSP framework and DWD warnings was 76% across all grid cells. Forested and agricultural areas consistently showed the highest similarities (82.6%), with similarities reaching up to 86% in some rural locations. In contrast, there was significant lower similarity in urban areas (73.9%), ranging from 59% to 78%. Industrial areas

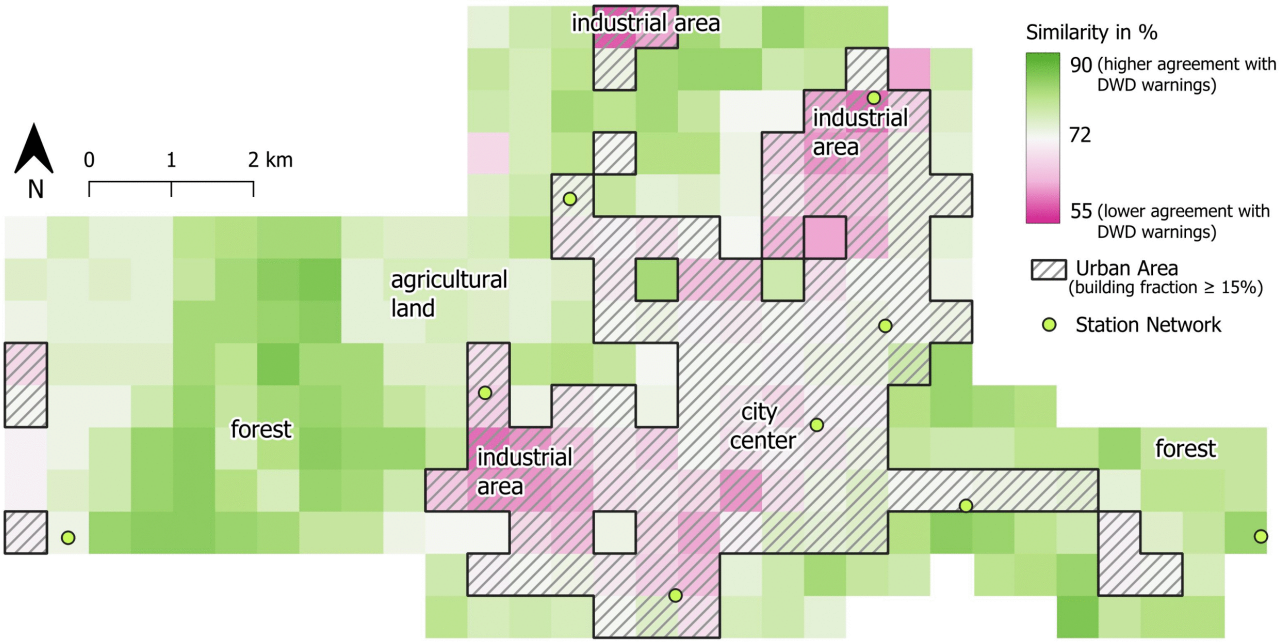

Fig 8. Map of study area showing the overall heat warning similarity in %. Values represent the percentage when PHSP warnings matched the official DWD warning level at each 512 x 512 m pixel (JJA 2023).

https://doi.org/10.1371/journal.pclm.0000941.g008

showed the lowest similarity, since these areas frequently received PHSP warnings despite no official DWD county-scale warning was issued. For the city center and dense residential areas intermediate similarities (65–75%) were observed.

Fig 9 delineates the spatial distribution of the PHSP's positive warning bias, quantified as the percentage of overestimation relative to official DWD warnings. PHSP demonstrated strong fidelity in the surrounding natural landscapes, with near-zero warning bias and high agreement with DWD warnings observed throughout the forested area. In contrast, a pronounced and systematic warning bias was evident within the urban areas, where the frequency of predicting a higher warning level reached up to 50%, particularly within the city center and industrial zones. For the whole study extent, the average warning bias was 22.5%.

As the choice of cell size is of major importance for neighborhood-scale warning, a comparison of the similarities and warning bias results in respect to cell size are shown in Table 5. The averaged overall similarity ranged from 77.7% to 78.9% for rural areas and 69.3% to 71.6% for urban areas, moving from the smallest to the largest grid size. This represents a maximum variation of only 1.2% (rural) and 2.3% (urban) across the tested resolutions. These small margins indicate that while increasing the grid cell size leads to a slight increase in agreement with the official DWD county warnings, the impact of spatial resolution is minimal. Consistently, the similarity was higher for rural areas compared to urban areas across all resolutions. For the case of PHSP issuing a warning of one or two level higher than the official DWD warning, the averaged warning bias ranged from 15.7% to 16.5% with no linear increase from highest to lowest resolution in rural areas. In urban areas the warning bias even decreased by 2% from highest resolution (30%) to lowest resolution (28%). Overall, the difference in similarity and warning bias were nearly resolution-independent between the scales studied (128 m - 1024 m).

The pattern seen in Figs 8 and 9 were likely linked to the morphological and land cover characteristics of the grid cells. With highest warning bias and lowest similarity in areas characterized by high building fraction, extensive sealed surfaces, and limited vegetation. Fig 10 shows the overall similarity (a) and warning bias (b) of the PHSP framework as functions

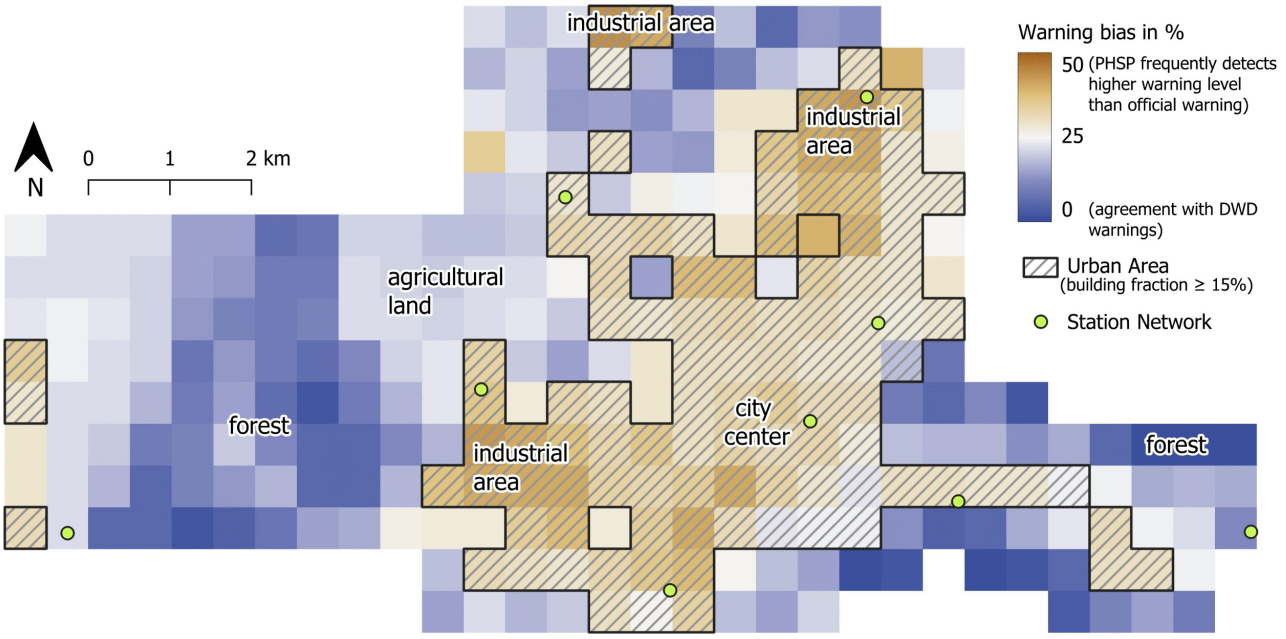

Fig 9. Map of study area showing the spatial distribution of PHSP warning bias compared to official heat warnings in %. Values represent the percentage of days when PHSP detected at least one warning level higher the official heat warnings at each 512 x 512 m pixel (JJA 2023).

https://doi.org/10.1371/journal.pclm.0000941.g009

Table 5. Average overall similarity and PHSP warning bias regarding official DWD heat warnings for different grid cell sizes separated in rural and urban areas.

Grid cell size	Rural areas		Urban areas	
	Similarity [%]	PHSP Warning bias[%]	Similarity [%]	PHSP Warning bias[%]
1024 m x 1024 m	78.9	16.5	71.6	28.0
512 m x 512 m	78.8	15.9	70.1	29.7
256 m x 256 m	78.3	15.7	69.6	30.0
128 m x 128 m	77.7	16.2	69.3	30.0

https://doi.org/10.1371/journal.pclm.0000941.t005

of the building fraction. A linear regression of the overall similarity of the PHSP framework compared to the official DWD warning product as a function of the building fraction revealed a highly significant negative relationship ($p > 0.05$, $r > 0.7$). With an increase in building fraction by 1%, the overall similarity declines by 2.6%. The regression between the warning bias of PHSP framework compared to the DWD warning product revealed a highly significant positive linear relationship ($p > 0.05$, $r > 0.7$), indicating that with increased building fraction by 1%, the warning bias increases as well by 3.3%.

Discussion

Validation and performance of the PHSP framework

The neighborhood-scale PHSP framework demonstrated varied but reasonable accuracy/similarity when evaluated against observational data and compared to official heat warnings. When compared against the street-level weather station network, the PHSP framework showed a clear performance difference between urban and rural settings. It achieved robust accuracy for urban stations, with an average of 82.0% and an average macro F1 score of 0.82, whereas

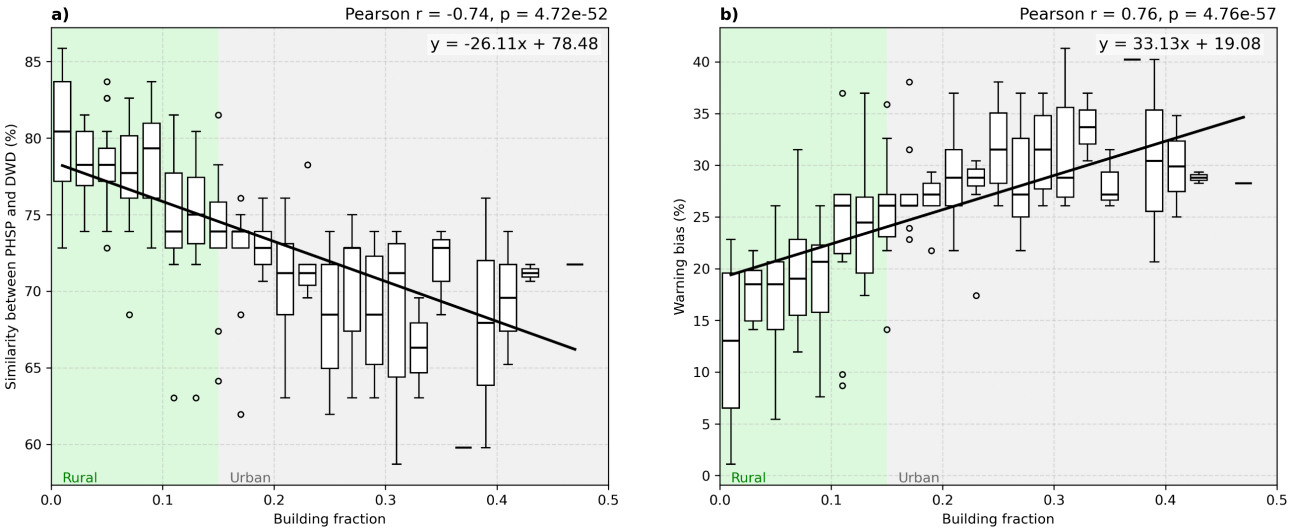

Fig 10. a) Relationship between overall similarity (PHSP vs DWD in %) and averaged building fraction of the 512 x 512 m cells and b) Relationship between warning bias of the DWD level by PHSP by at least 1 warning level and averaged building fraction of the 512 x 512 m cells. Boxplots show the interquartile range (IQR, 25–75%) of values per 0.02 building-fraction bin. Whiskers extend to 1.5 × IQR, points beyond are outliers. A linear regression is indicated with the black line.

https://doi.org/10.1371/journal.pclm.0000941.g010

performance in rural areas was lower, with an average accuracy of 75.6% and a macro F1 score of 0.70 (Table 3). It suggests that the grid cells in urban settings capture the reality given by the weather station network better, than the rural grid cells, even though urban areas are characterized by more heterogenous surface characteristics. However, it should be noted that the PHSP was compared to only two rural stations, one of which was particularly underperforming, possibly skewing the results. Additionally, the method of validating gridded predictions against point observations poses a problem in itself, as the representativeness of point data decreases with increasing grid cell size. The method further assumes that the stations are situated at locations representative of the surrounding area. While this was a key design criterion during installation [48] Plein et al., this assumption may hold true to varying degrees.

Official DWD warnings and the PHSP warnings have overall similarities of 70.1% in urban areas and 78.8% in rural areas (Fig 8) indicating a generally high agreement. However, the distinct discrepancies in False Alarm Rates (FAR), recall, and temporal onset reveal significant differences in how local microclimates versus regional warnings account for heat stress. The systematic underperformance in detecting L1 warnings in both urban (F1 score = 0.35) and rural (F1 score = 0.42) contexts suggests that the UTCI threshold chosen in the PHSP framework is more sensitive and might not fully compare with the official warning criteria (Table 4). This sensitivity resulted in a high number of “false” positives relative to the official DWD warning, specifically in urban areas where the PHSP triggered an L1 warning on 24 days when the DWD issued no warning at all. For L2, the PHSP behaved differently across contexts: In urban areas the PHSP showed a nearly perfect recall, ensuring no official heat wave was missed, though even issuing warnings on days the DWD did not. In rural areas, the system was more conservative (recall = 0.67), failing to detect some official warning days, but maintaining high reliability (precision = 0.76) when it did warn.

Hypothesis 1: Neighborhood-scale spatial heterogeneity

The results strongly support the hypothesis that neighborhood-scale heat stress exhibits substantial spatial heterogeneity within cities and between urban and rural areas that isn’t captured by the current county-level system of DWD. The distribution of heat warning days for urban and rural grid cells shows a structural difference in heat exposure (Fig 5). Urban

areas are characterized by frequent and more uniform heat stress, whereas rural areas exhibit a high variability. In this study, the intraurban variability was lower than that between urban and rural areas (Fig 5). The analysis against the official heat warning products revealed distinct geographic patterns in warning similarity, with forested areas achieving high agreement (up to 86%) while built-up areas showed significantly lower agreement (59% to 78%) (Fig 8). Industrial zones and the city center showed intermediate to poor agreement (65–75%), reflecting the complex three-dimensional urban structure that drives high spatial variability in thermal comfort [31].

This heterogeneity is systematically linked to urban morphology and land cover. The strong negative linear relationship between building fraction and warning similarity (Fig 10a), where similarity declines by 2.6% for every 1% increase in building fraction, confirms that urban form is a key driver of model disagreement. The identification of systematic warning bias in built-up areas, reaching up to 50% of days in the city center and industrial zones (Fig 9), demonstrates that different urban morphologies require differentiated approaches to heat risk assessment. This spatial pattern confirms that the discrepancies between PHSP and official warnings reflect genuine, microclimatic differences in heat stress conditions rather than methodological artifacts.

Hypothesis 2: Urban heat amplification

The second hypothesis, that urban areas demonstrate more frequent and intense heat stress, is confirmed by the quantitative analysis. The analysis of warning frequency both for station data and neighborhood averaged PHSP data reveals a clear intensification of heat stress in urban environments.

Regarding warning day counts, the official DWD system issued 21 warning days, whereas urban grid cells in the PHSP framework experienced up to 55 L1 warning days, with 94% of all cells in the study area exceeding the official count (Fig 4).

The PHSP detected significantly more heat warning days on average in urban areas (27.4 L1, 17.5 L2) than in rural areas (22.7 L1, 9.6 L2). This translates to urban areas experiencing 1.2 times the frequency of L1 incidents and 1.8 times the frequency of L2 incidents compared to rural areas, providing clear evidence of urban effects on human thermal comfort.

The temporal analysis further underscores this, revealing that the first PHSP L1 warning of the year in urban areas occurred 18 days earlier, and the first L 2 warning 19 days earlier, than the official DWD warnings. In contrast, rural areas showed much closer alignment (2 days and 0 days earlier, respectively), highlighting the inadequacy of county-scale systems for capturing early-season urban heat stress. This suggests that while the DWD warnings, which cover the entire Breisgau-Hochschwarzwald county, accurately reflect the general rural baseline, they may mask the accelerated onset of heat seasons in urban centers.

Hypothesis 3: Systematic biases in county-scale warnings

The results provide strong evidence for the third hypothesis: that county-scale warnings exhibit systematic biases and under-detect heat in urban neighborhoods. The comparison revealed that official DWD warnings failed to capture a substantial number of heat stress days identified by the PHSP in urban areas. Over the study period, the PHSP flagged 24 days requiring a warning (20.8 L1, 3.2 L2) in urban settings when the DWD issued no warning at all. This systematic under-detection is a critical finding, especially in industrial and dense residential areas (Figs 8 and 9). The superior agreement and lower warning bias in rural areas (78.8% similarity) compared to urban areas (70.1% similarity) confirms that current county-scale systems are better calibrated for non-urban environments. This bias disproportionately affects high-risk urban neighborhoods, where the combination of dense infrastructure and vulnerable populations can exacerbate heat-related health impacts.

Advantages of physiologically-based neighborhood-scale approaches

The physiologically-based PHSP approach offers several advantages over existing air temperature-based systems. By incorporating radiation, humidity, and wind at pedestrian-level through the UTCI, it captures crucial aspects of thermal

comfort in complex urban environments that are missed by temperature-only metrics [23]. The use of UTCI thresholds linked to documented physiological responses provides objective, health-relevant criteria for heat warnings. Critically, the physiological responses underpinning the PHSP threshold of L1 and L2 correspond to documented population-level health outcomes [42–46], such as sustained dehydration and progressive cardiovascular strain, leading to increased cardiovascular mortality by 15–24% and respiratory mortality by 34% [14,15]. By incorporating duration requirements (i.e., h_{32} and h_{38}, see Table 1), the PHSP provides a more comprehensive assessment of human heat stress by more accurately reflecting the mechanisms through which heat translates into adverse health outcomes than systems that are triggered by instantaneous temperature peaks. This is in alignment with findings from other studies implementing advanced heat exposure measures [65].

The neighborhood-scale resolution (512 × 512 m) enables the identification of intra-urban heat hotspots, providing actionable information for targeted public health interventions. Critical infrastructure within a neighborhood-scale grid cell, such as hospitals, can be warned and informed more targeted to better regulate indoor heat stress [66].

This spatial specificity allows for more effective use of limited resources, as demonstrated in similar applications [22,35,67]. As shown in Table 5, the methodology is not restricted to this grid size and could be adapted to other relevant spatial units, such as administrative city districts, to enhance communication and action.

Implications for public health and heat warning systems

To our knowledge, this study is the first to directly compare neighborhood-scale, physiologically-based outdoor heat stress predictions with official county-scale warnings. The results reveal that current warning systems cannot account for urban-rural and intra-urban variability, due to their nature of county-scale warnings. This gap represents a public health concern, especially as urban populations grow and climate change intensifies heat events. The finding is concerning given that other studies have found that official warnings in Germany already miss a significant portion of societal heat responses [30]. The 24 urban warning days identified by PHSP but missed by the county-level system therefore represent days on which no information and instructions to take protective action reached the urban public. Our study additionally adds a spatial dimension to that concern: even on warned days, the county-level system cannot communicate that urban neighborhoods experience more severe conditions than the regional average.

The higher frequency of very strong heat in urban areas (3.2× for Level 2 conditions) suggests that climate impacts will be disproportionately severe in cities. With over 75% of the German population living in cities [68], our results highlight a critical need to adjust warning criteria and warning systems to better reflect actual urban heat vulnerability. A county-level warning in the case of Freiburg covering forests, agricultural land, and dense urban neighborhoods simultaneously cannot fulfill this function for the urban populations. Transitioning to neighborhood-scale systems would not only improve warning accuracy but enable differentiated public health responses: targeted outreach to vulnerable residents, activation of local cooling strategies and real-time guidance for healthcare providers in affected districts [21,22,25].

Several important directions for future HHWSs emerge from this study. Currently, the framework neglects variability of urban characteristics within a grid cell. Thus, a measure of heterogeneity within the gird cell to account for this should be added. Importantly, the PHSP as developed here allows only for a hazard warning: it identifies where and when outdoor thermal conditions exceed physiologically critical thresholds, but it does not yet take exposure and susceptibility into account of urban residents. Translating hazard into risk requires two additional components: exposure and vulnerability. Heat risk can be assessed by accounting for the presence, duration, and activity patterns of urban populations in affected outdoor spaces, which vary substantially by time of day, land use, and season, as well as by the different susceptibility of population subgroups divided by age, socioeconomic status, pre-existing health conditions, access to preventive infrastructure and urban green [22,35]. The PHSP framework provides the necessary physiologically-grounded hazard product and by incorporating exposure and vulnerability through intersection of appropriate data sets (e.g., gridded census data), transition toward a fully integrated heat risk warning system would be possible.

Furthermore, systematic comparisons across multiple cities are needed to enable the development of more robust and transferable methodologies. Operationally, it needs to be established whether the heightened warning frequency of the PHSP framework causes a warning fatigue effect. Finally, longitudinal health outcome studies are necessary to evaluate the public health relevance of neighborhood-scale predictions by correlating them with heat-related morbidity and mortality data [13,26,50].

Methodological considerations and limitations

A sensitivity analysis demonstrated that the choice of grid cell size (from 128 m to 1024 m) had only a marginal impact on similarity and warning bias, suggesting the results are resolution-independent within this range (Table 5) and could also be applied to administrative units of roughly same scale. However, this resolution necessarily smooths fine-scale thermal variations important for individual exposure. Currently, the approach is still requiring the calculation of UTCI at 1 x 1 m and hourly resolution. The spatial aggregation to a 512 × 512 m resolution is a reasonable compromise between the potential of computational efficiency and neighborhood-scale applicability, comparable to other high-resolution prototypes [35,67].

The focus on daytime hours aligns with pedestrian activity but cannot account for nighttime heat stress, which is critical for vulnerable populations, in particular the associated indoor heat stress [66]. The duration thresholds (2 hours for L1, 0.5 hours for L2) are based on physiological studies but may require regional adaptation based on local acclimatization patterns [30,69]. The HTC-NN framework, while efficient, does not account for certain meso-scale processes like thermo-orographic wind systems, as all data shown here was driven by a single model grid cell from the ICON-D2 forecast [31] by only considering urban morphometry and land cover.

The choice of UTCI over the Perceived Temperature (PT) used in the official DWD system deserves consideration. While the DWD system is specifically calibrated for the German population [21], UTCI offers a universal framework applicable across different climates. However, the poor performance of PHSP in detecting official L1 warnings suggests that a direct comparison between UTCI and PT requires careful calibration.

The HTC-NN framework enables rapid generation of high-resolution heat stress predictions suitable for daily operational use while maintaining physical realism through training on comprehensive numerical simulations [31]. However, the machine learning approach inherits any biases present in the training data, and model performance may degrade when applied to meteorological conditions or urban morphologies not well-represented in the training dataset [31]. The model does not account for the influence of surrounding terrain or topographic features such as thermo-orographic wind systems [31]. Using meteorological data from several NWP models (e.g., ICON D2 grid cells) for forcing could account for larger mesoscale processes.

Transferability

The methodological framework presented in our study is designed to be transferable to other urban contexts. The HTC-NN can be retrained for different urban morphologies and climates given the geospatial and meteorological input data is available [11,31]. The PHSP aggregation is not bound to any Freiburg-specific parameter. Additionally, the physiological basis of the UTCI is designed to be universally applicable [18].

However, the specific findings reported, including warning frequencies, urban-rural disparities and the magnitude of county-scale under-detection, are tied to the particular characteristics of Freiburg. Even if the results might be similar for other central European cities with comparable urban characteristics, these insights should not be assumed to be directly translatable to cities with different geographical, environmental or social conditions. The UTCI thresholds applied here may require recalibration to reflect local heat tolerance, as populations in hotter climates may tolerate higher UTCI values, while those in cooler regions may experience heat stress at lower thresholds [30,69]. While the UTCI is calibrated to an active average adult and doesn't capture the elevated risk of vulnerable groups [18], the PHSP framework allows for lowering the UTCI threshold values as well as the exceedance durations. This permits to issue earlier or more

frequent warnings for population groups with elevated heat sensitivity without any structural changes to the framework. In principle, the same system could run in parallel at different sensitivity levels for general and vulnerable populations respectively.

Conclusions

This study demonstrated the feasibility and value of a neighborhood-scale, physiologically-based heat warning framework (PHSP) for a case study in Freiburg, Germany during the summer of 2023. The key findings reveal limitations in current county-scale systems and highlight the advantages of a high-resolution approach to model heat.

First, the research confirms systematic urban-rural disparities in heat stress that current official warnings at county-level are not able to capture. The PHSP identified 24 warning days in urban areas that were completely missed by the county-level system, pointing to a significant under-detection of heat. Second, this study provides clear, quantified evidence of urban heat amplification: urban areas experienced 1.2 times the frequency of strong (L1) and 1.8 times the frequency of very strong (L2) heat incidents than surrounding rural areas. Third, the results show pronounced intra-urban heterogeneity in heat risk that is directly linked to urban morphology and land cover. Consequently, industrial zones and dense city centers remain most affected by heat and least protected by current county-level warnings. Finally, the methodological validation confirms the framework's robustness, achieving reasonable accuracy against both official warnings (70.1% urban, 78.8% rural) and on-the-ground weather station observations (82.0% urban, 75.6% rural).

The machine-learning-based PHSP framework demonstrates a computationally efficient and transferable means to address the inherent limitations of county-level warning systems. Collectively, the results emphasize the necessity of evolving from coarse approaches to a spatially refined, next-generation heat-health warning system.

Acknowledgments

We gratefully acknowledge Matthias Zeeman from the University of Freiburg for the data management of the weather station network and staff of the University of Freiburg for establishment and operation of the network.

Author contributions

Conceptualization: Svenja Ludwig, Andreas Christen.

Data curation: Svenja Ludwig, Ferdinand Briegel.

Formal analysis: Svenja Ludwig.

Funding acquisition: Andreas Christen.

Methodology: Svenja Ludwig.

Project administration: Andreas Christen.

Resources: Andreas Christen.

Software: Ferdinand Briegel.

Supervision: Ferdinand Briegel, Andreas Christen.

Validation: Svenja Ludwig.

Visualization: Svenja Ludwig.

Writing – original draft: Svenja Ludwig.

Writing – review & editing: Ferdinand Briegel, Andreas Christen.

References

1. Intergovernmental Panel on Climate Change IPCC. Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. 1st ed. Cambridge University Press; 2023. https://doi.org/10.1017/9781009157896
2. Winklmayer C, Matthies-Wiesler F, Muthers S, Buchien S, Kuch B, An Der Heiden M. Heat in Germany: Health risks and preventive measures. 2023. https://doi.org/10.25646/11651
3. Kuttler W. Stadtklima. UWSF - Z Umweltchem Ökotox. 2004;16(3):187–99. https://doi.org/10.1065/uwsf2004.03.078
4. Oke TR. Urban climates. Cambridge: Cambridge University Press; 2017.
5. Kuttler W. Stadtklima. UWSF - Z Umweltchem Ökotox. 2004;16(4):263–74. https://doi.org/10.1065/uwsf2004.08.083
6. Wösle J, Briegel F, Zeeman M, Plein M, Christen A, Matzarakis A. Intra-urbane Variabilität der Intensität und Häufigkeit von Hitzebelastung im Stadtgebiet von Freiburg. Gefahrstoffe Reinhaltung der Luft. 2025;85(7–8):175–82.
7. Ng E, Yuan C, Chen L, Ren C, Fung JCH. Improving the wind environment in high-density cities by understanding urban morphology and surface roughness: a study in Hong Kong. Landsc Urban Plan. 2011;101(1):59–74. https://doi.org/10.1016/j.landurbplan.2011.01.004 PMID: 32287617
8. Ketterer C, Gangwisch M, Fröhlich D, Matzarakis A. Comparison of selected approaches for urban roughness determination based on voronoi cells. Int J Biometeorol. 2017;61(1):189–98. https://doi.org/10.1007/s00484-016-1203-2 PMID: 27369974
9. Matzarakis A, Rutz F, Mayer H. Modelling radiation fluxes in simple and complex environments—application of the RayMan model. Int J Biometeorol. 2007;51(4):323–34. https://doi.org/10.1007/s00484-006-0061-8
10. Martinelli L, Lin T-P, Matzarakis A. Assessment of the influence of daily shadings pattern on human thermal comfort and attendance in Rome during summer period. Build Environ. 2015;92:30–8. https://doi.org/10.1016/j.buildenv.2015.04.013
11. Briegel F, Makansi O, Brox T, Matzarakis A, Christen A. Modelling long-term thermal comfort conditions in urban environments using a deep convolutional encoder-decoder as a computational shortcut. Urban Climate. 2023;47:101359. https://doi.org/10.1016/j.uclim.2022.101359
12. Mücke HG, Litvinovitch JM. Heat extremes, public health impacts, and adaptation policy in Germany. Int J Environ Res Public Health. 2020;17(21):7862. https://doi.org/10.3390/ijerph17217862
13. Wang J, Nikolaou N, An der Heiden M, Irrgang C. High-resolution modeling and projection of heat-related mortality in Germany under climate change. Commun Med (Lond). 2024;4(1):206. https://doi.org/10.1038/s43856-024-00643-3 PMID: 39433977
14. Zhang S, Breitner S, De' Donato F, Stafoggia M, Nikolaou N, Aunan K, et al. Heat and cause-specific cardiopulmonary mortality in Germany: a case-crossover study using small-area assessment. Lancet Reg Health Eur. 2024;46:101049. https://doi.org/10.1016/j.lanepe.2024.101049 PMID: 39290807
15. Zacharias S, Koppe C, Mücke HG. Influence of heat waves on ischemic heart diseases in Germany. Climate. 2014;2(3):133–52. https://doi.org/10.3390/cli2030133
16. Matzarakis A, Nouri AS. Perception of heat stress in cities and measures for health protection. PLOS Clim. 2022;1(11):e0000104. https://doi.org/10.1371/journal.pclm.0000104
17. Blazejczyk K, Epstein Y, Jendritzky G, Staiger H, Tinz B. Comparison of UTCI to selected thermal indices. Int J Biometeorol. 2012;56(3):515–35. https://doi.org/10.1007/s00484-011-0453-2 PMID: 21614619
18. Błażejczyk K, Jendritzky G, Bröde P, Fiala D, Havenith G, Epstein Y. An introduction to the Universal Thermal Climate Index (UTCI). Geogr Pol. 2013;86(1):5–10. https://doi.org/10.7163/GPol.2013.1
19. Pappenberger F, Jendritzky G, Staiger H, Dutra E, Di Giuseppe F, Richardson DS, et al. Global forecasting of thermal health hazards: the skill of probabilistic predictions of the Universal Thermal Climate Index (UTCI). Int J Biometeorol. 2015;59(3):311–23. https://doi.org/10.1007/s00484-014-0843-3 PMID: 24859883
20. Di Napoli C, Pappenberger F, Cloke HL. Verification of heat stress thresholds for a health-based heat-wave definition. J Appl Meteorol Climatol. 2019;58(6):1177–94. https://doi.org/10.1175/jamc-d-18-0246.1
21. Matzarakis A, Laschewski G, Muthers S. The heat health warning system in Germany—application and warnings for 2005 to 2019. Atmosphere. 2020;11(2):170. https://doi.org/10.3390/atmos11020170
22. Gangwisch M, Matzarakis A. Composition of factors for local heat adaptation measures at the local level in cities of the mid-latitude - An approach for the south-west of Germany. Environ Int. 2024;187:108718. https://doi.org/10.1016/j.envint.2024.108718 PMID: 38735079
23. Casanueva A, Burgstall A, Kotlarski S, Messeri A, Morabito M, Flouris AD. Overview of Existing Heat-Health Warning Systems in Europe. Int J Environ Res Public Health. 2019;16(15):2657. https://doi.org/10.3390/ijerph16152657
24. Kotharkar R, Ghosh A. Progress in extreme heat management and warning systems: a systematic review of heat-health action plans (1995-2020). Sustain Cities Soc. 2022;76:103487. https://doi.org/10.1016/j.scs.2021.103487
25. Matzarakis A, Giannaros C. Toward the next-generation of heat-health warning systems and action plans. Atmosphere. 2025;16(8):938. https://doi.org/10.3390/atmos16080938
26. Ghada W, Estrella N, Ankerst DP, Menzel A. Universal thermal climate index associations with mortality, hospital admissions, and road accidents in Bavaria. PLoS One. 2021;16(11):e0259086. https://doi.org/10.1371/journal.pone.0259086 PMID: 34788302

27. Giannaros C, Economou T, Parliari D, Galanaki E, Kotroni V, Lagouvardos K, et al. A thermo-physiologically consistent approach for studying the heat-health nexus with hierarchical generalized additive modelling: Application in Athens urban area (Greece). Urban Clim. 2024;58:102206. https://doi.org/10.1016/j.uclim.2024.102206
28. Golechha M, Shah P, Mavalankar D. Threshold determination and temperature trends analysis of Indian cities for effective implementation of an early warning system. Urban Clim. 2021;39:100934. https://doi.org/10.1016/j.uclim.2021.100934
29. Rizmie D, Christo E, Carosella EA, Almstead L, Bhinge M, Sivasankaran A. The role of climate services for health: theoretical case studies on heat-health warning systems in India. PLOS Clim. 2025;4(12):e0000637. https://doi.org/10.1371/journal.pclm.0000637
30. Bogdanovich E, Brenning A, Reichstein M, De Polt K, Guenther L, Frank D, et al. Official heat warnings miss situations with a detectable societal heat response in European countries. Int J Disas Risk Reduc. 2024;100:104206. https://doi.org/10.1016/j.ijdrr.2023.104206
31. Briegel F, Wehrle J, Schindler D, Christen A. High-resolution multi-scaling of outdoor human thermal comfort and its intra-urban variability based on machine learning. Geosci Model Dev. 2024;17(4):1667–88. https://doi.org/10.5194/gmd-17-1667-2024
32. Veisi O, Tehrani AA, Gharaei B, Du DK, Shakibamanesh A. Towards universal thermal climate index prediction via machine learning approaches. Renew Sustain Energy Rev. 2025;217:115680. https://doi.org/10.1016/j.rser.2025.115680
33. Zhang X, Chen T, Zhang S, Yuan C, Tan PY. Toward healthy neighborhood thermal environments: Interpretation of thermal comfort and rapid identification of hotspot areas using image-based metrics and explainable machine learning. Sustain Cities Soc. 2025;130:106648. https://doi.org/10.1016/j.scs.2025.106648
34. Zheng L, Lu W. Urban micro-scale street thermal comfort prediction using a 'graph attention network' model. Build Environ. 2024;262:111780. https://doi.org/10.1016/j.buildenv.2024.111780
35. Kacker K, Srivastava P, Mukherjee M. Personalized heat stress early warning system for an urban area. Environ Int. 2025;199:109507. https://doi.org/10.1016/j.envint.2025.109507 PMID: 40328089
36. Briegel F, Schrodi S, Sulzer M, Brox T, Pinto JG, Christen A. Deep learning enables city-wide climate projections of street-level heat stress. Urban Clim. 2025;62:102564. https://doi.org/10.1016/j.uclim.2025.102564
37. Briegel F, Pinto JG, Christen A. Is satellite land surface temperature an appropriate proxy for intra-urban variability of daytime heat stress? Remote Sens Environ. 2025;331:115045. https://doi.org/10.1016/j.rse.2025.115045
38. Fünfgeld H, Christen A, Briegel F, Schrodi S, Speidel A, Felder C, et al. Optimizing urban greening and densification in the context of outdoor heat: opportunities for AI-supported urban adaptation. Landsc Urban Plann. 2026;268:105574. https://doi.org/10.1016/j.landurbplan.2025.105574
39. Reinert D, Prill F, Frank H, Denhard M, Baldauf M, Schraff C, et al. DWD Database Reference for the Global and Regional ICON and ICON-EPS Forecasting System. 2025.
40. Deutscher W. Operational NWP system: ICON-D2/ ICON-D2-EPS/ RUC [Internet]. 2025. Available from: https://www.dwd.de/DE/fachnutzer/forschung_lehre/numerische_wettervorhersage/numerische_wettervorhersage_node.html
41. Wehrle J, Jung C, Giometto M, Christen A, Schindler D. Introducing new morphometric parameters to improve urban canopy air flow modeling: A CFD to machine-learning study in real urban environments. Urban Clim. 2024;58:102173. https://doi.org/10.1016/j.uclim.2024.102173
42. Sawka MN, Burke LM, Eichner ER, Maughan RJ, Montain S, Stachenfeld N. Exercise and fluid replacement. Med Sci Sports Exerc. 2007;39(2):377–90. https://doi.org/10.1249/mss.0b013e31802ca597
43. Glaser J, Lemery J, Rajagopalan B, Diaz HF, García-Trabanino R, Taduri G, et al. Climate change and the emergent epidemic of CKD from heat stress in rural communities: the case for heat stress nephropathy. Clin J Am Soc Nephrol. 2016;11(8):1472–83. https://doi.org/10.2215/CJN.13841215 PMID: 27151892
44. Bouchama A, Knochel JP. Heat stroke. N Engl J Med. 2002;346(25):1978–88. https://doi.org/10.1056/NEJMra011089
45. Crandall CG, González-Alonso J. Cardiovascular function in the heat-stressed human. Acta Physiol (Oxf). 2010;199(4):407–23. https://doi.org/10.1111/j.1748-1716.2010.02119.x PMID: 20345414
46. Ebi KL, Capon A, Berry P, Broderick C, de Dear R, Havenith G, et al. Hot weather and heat extremes: health risks. Lancet. 2021;398(10301):698–708. https://doi.org/10.1016/S0140-6736(21)01208-3 PMID: 34419205
47. Briegel F, Christen A. High-resolution land cover classification for Freiburg im Breisgau (Germany) [Internet]. University of Freiburg; 2025 [cited 2026 Jan 19]. Available from: https://zenodo.org/doi/10.5281/zenodo.14855706
48. Plein M, Kersten F, Zeeman M, Christen A. Street-level weather station network in Freiburg, Germany: Station documentation [Internet]. Zenodo; 2024 [cited 2026 Jan 19]. Available from: https://zenodo.org/doi/10.5281/zenodo.12732551
49. Statistisches Landesamt Baden-Württemberg. Bevölkerung nach Nationalität – vierteljährlich (Basis Zensus 2022). [Internet]. 2022. Available from: https://www.statistik-bw.de
50. Muthers S, Laschewski G, Matzarakis A. The summers 2003 and 2015 in South-West Germany: heat waves and heat-related mortality in the context of climate change. Atmosphere. 2017;8(11):224. https://doi.org/10.3390/atmos8110224
51. Plein M, Feigel G, Zeeman M, Dormann CF, Christen A. Using gradient boosting for gap-filling to analyze temperature and humidity patterns in an urban weather station network in Freiburg, Germany. Urban Clim. 2025;62:102496. https://doi.org/10.1016/j.uclim.2025.102496
52. Stewart ID, Oke TR. Local climate zones for urban temperature studies. Bull Am Meteorol Soc. 2012;93(12):1879–900. https://doi.org/10.1175/bams-d-11-00019.1

53. Demuzere M, Kittner J, Martilli A, Mills G, Moede C, Stewart ID, et al. A global map of local climate zones to support earth system modelling and urban-scale environmental science. Earth Syst Sci Data. 2022;14(8):3835–73. https://doi.org/10.5194/essd-14-3835-2022
54. Feigel G, Plein M, Zeeman M, Metzger S, Matzarakis A, Schindler D, et al. High spatio-temporal and continuous monitoring of outdoor thermal comfort in urban areas: a generic and modular sensor network and outreach platform. Sustain Cities Soc. 2025;119:105991. https://doi.org/10.1016/j.scs.2024.105991
55. Deutscher Wetterdienst. Monthly station observations for station 1443 [Internet]. 2025. Available from: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/monthly/kl/historical/
56. Zeeman M, Christen A, Grimmond S, Fenner D, Morrison W, Feigel G, et al. Modular approach to near-time data management for multi-city atmospheric environmental observation campaigns. Geosci Instrum Method Data Syst. 2024;13(2):393–424. https://doi.org/10.5194/gi-13-393-2024
57. VDI. Environmental meteorology—Methods for human biometeorological evaluation of the thermal component of the climate. Beuth Verlag; 2022.
58. Campbell Scientific Inc. Black globe temperature sensor: Measuring heat stress. [Internet]. 2024. Available from: https://www.campbellsci.com/blackglobe
59. Briegel. Code HTC-NN [Internet]. Zenodo; 2023 [cited 2026 Jan 19]. Available from: https://zenodo.org/record/7974472
60. Opitz J. From Bias and Prevalence to Macro F1, Kappa, and MCC: A structured overview of metrics for multi-class evaluation. 2022. Available from: https://api.semanticscholar.org/CorpusID:253270558
61. Staiger H, Laschewski G, Grätz A. The perceived temperature - a versatile index for the assessment of the human thermal environment. Part A: scientific basics. Int J Biometeorol. 2012;56(1):165–76. https://doi.org/10.1007/s00484-011-0409-6 PMID: 21336880
62. Pfafferott J, Becker P. Erweiterung des Hitzewarnsystems um die Vorhersage der Wärmebelastung in Innenräumen. Bauphysik. 2008;30(4):237–43. https://doi.org/10.1002/bapi.200810031
63. Deutscher Wetterdienst. Historical heat warnings [Internet]. 2025. Available from: https://opendata.dwd.de/climate_environment/health/historical_alerts/heat_warnings/
64. Statistisches Landesamt Baden-Württemberg. Flächenerhebung nach Art der tatsächlichen Nutzung 2023. 2024.
65. Sadeghi M, de Dear R, Morgan G, Santamouris M, Jalaludin B. Development of a heat stress exposure metric – Impact of intensity and duration of exposure to heat on physiological thermal regulation. Build Environ. 2021;200:107947. https://doi.org/10.1016/j.buildenv.2021.107947
66. Epp K, Sulzer M, Steinmann D, Zeeman M, Matzarakis A, Christen A. Assessment of current and future heat in a large hospital complex based on continuous indoor measurements and climate simulations. Atmosph Meteorol Climatol Haz. 2025. https://doi.org/10.5194/egusphere-2025-3871
67. Giannaros C, Galanaki E, Agathangelidis I. HEAT-ALARM: Heat–health warning system manual [Internet]. 2024. Available from: https://osf.io/vy84u
68. United Nations. World urbanization prospects: the 2018 revision. New York: United Nations; 2019.
69. Pascal M, Wagner V, Le Tertre A, Laaidi K, Honoré C, Bénichou F, et al. Definition of temperature thresholds: the example of the French heat wave warning system. Int J Biometeorol. 2013;57(1):21–9. https://doi.org/10.1007/s00484-012-0530-1 PMID: 22361805