High-resolution multi-scaling of outdoor human thermal comfort and its intra-urban variability based on machine learning

Ferdinand Briegel[1], Jonas Wehrle[1], Dirk Schindler[1], Andreas Christen[1]

[1]Chair of Environmental Meteorology, Faculty of Environment and Natural Resources, University of Freiburg, Freiburg im Breisgau, Germany

Corresponding author: Ferdinand Briegel (ferdinand.briegel@meteo.uni-freiburg.de)

Abstract. As the frequency and intensity of heat waves will continue to increase in the future, accurate and high-resolution mapping and forecasting of human outdoor thermal comfort in urban environments is of great importance. This study presents a machine learning based outdoor thermal comfort model with a good trade-off between computational cost, complexity, and accuracy compared to common numerical urban climate models. The machine learning approach is basically an emulation of different numerical urban climate models. The final model consists of four sub-models that predict air temperature, relative humidity, wind speed, and mean radiant temperature based on meteorological forcing and geospatial data on building form, land cover, and vegetation. These variables are then combined into a thermal index (Universal Thermal Climate Index - UTCI). All four sub-model predictions and the final model output are evaluated using street-level measurements from a dense urban sensor network in Freiburg, Germany. The final model has a mean absolute error of 2.3 K. Based on a city-wide simulation for the city of Freiburg we demonstrate that the model is fast and versatile enough to simulate multiple years at hourly timesteps to predict street-level UTCI at 1 m spatial resolution for an entire city. Simulations indicate that neighborhood-averaged thermal comfort conditions vary widely between neighborhoods, even if they are attributed to the same local climate zones, e.g. due to differences in age and degree of urban vegetation. Simulations also show contrasting differences in the location of hot spots during the day and at night.

1   Introduction

The frequency and severity of heatwaves have increased and are expected to increase even further due to human-caused climate change (IPCC, 2021). In addition, heatwaves are occurring earlier in summer, resulting in a longer period of potential heat stress (IPCC, 2021). Rousi et al. (2022) found that the frequency and intensity of heatwaves increased 3–4 times faster in Europe than in the rest of the mid-latitudes over the past decades. In 2020, heatwaves in Western Europe accounted for 42 % of all reported global deaths from extreme weather events, with a total of 6340 deaths (CRED, 2021). As the severity of heatwaves also depends on land cover and land use, urban areas are even more exposed to extreme heatwave events than rural areas due to their physical characteristics (Unger et al., 2020; Masson et al., 2020), including reduced nocturnal cooling and limited access to cool microenvironments for urban populations.

Human thermal comfort is influenced not only by air temperature (T_a) but also by wind speed (U), radiation, and humidity. The variables expressing the effect of radiation and humidity are the mean radiant temperature (T_{mrt}) and relative humidity (RH), respectively. Thermal indices combine these four environmental variables to describe thermal comfort and overall thermal stress of an individual (Epstein and Moran, 2006). Multiple thermal indices have been developed, such as the Physiological Equivalent Temperature (PET) or Universal Thermal Climate Index (UTCI) (Coccolo et al., 2016; Potchter et al., 2018). Several studies have concluded that T_{mrt} is the driver of outdoor human thermal comfort during daytime (Cohen et al., 2012; Holst and Mayer, 2011; Kántor and Unger, 2011). In addition to terrain, the complex and heterogeneous three-dimensional structure of cities causes high spatial and temporal variability of these environmental variables. T_{mrt} and U have the highest variability at the micro-scale (Matzarakis et al., 2016). However, T_a also varies locally, although less strongly (Fenner et al., 2017; Quanz et al., 2018; Shreevastava et al., 2021). In particular, T_a is more relevant to human thermal comfort during nighttime than T_{mrt} (Lee et al., 2013). As this study is concerned with a city-wide multi-scale approach to modeling outdoor human thermal comfort, the focus is on modeling T_a, RH, U, and T_{mrt} within the urban canopy layer (at about 1.1– 2.0 m a.g.l.) at the neighborhood scale (T_a and RH: 500 x 500 m) and building-resolved scale (T_{mrt} and U: 1 x 1 m).

In recent years, several deterministic and stochastic modeling approaches have been developed to map urban climate, outdoor thermal comfort, and canopy urban heat island (UHI) at different scales and layers and with varying complexity (Mirzaei, 2015). Mesoscale models have been parametrized for urban surfaces, the so-called slab or bulk models (Dupont et al., 2006), and coupled with urban canopy models (Chen et al.,2011; Hamdi et al. 2012; Martilli et al., 2002; Masson, 2000; Rafael et al., 2020). Urban canopy models, on the other hand, have also been used as stand-alone (offline) models to investigate the urban surface energy and water balance at the local scale (Best and Grimmond, 2015, and Grimmond et al., 2011). In addition to numerical urban climate models, statistical models (Chen et al., 2022; Ho et al., 2014; Straub et al., 2019) and dense measurement networks (Gubler et al., 2021) have been used to map urban climate. It can be concluded that there are several ways to model urban climate on various scales. However, as pointed out by Hamdi et al. (2020) and Masson et al. (2020), the complexity of the model should depend on the application and only be as complex as necessary.

Recently, machine and deep learning models (ML) have gained increasing attention in urban meteorology and have been used to emulate numerical urban climate models (Meyer et al., 2022; Briegel et al., 2023). The micro-scale T_{mrt} model SOLWEIG (SOlar and LongWave Environmental Irradiance Geometry – Lindberg and Grimmond, (2019)) was emulated by a deep convolutional encoder-decoder approach, called U-Net, which showed a promising trade-off between accuracy and computational cost (Briegel et al., 2023). Meyer et al. (2022) emulated an ensemble of Urban Land Surface Models (ULSM) using a simple Multi-Layer-Perceptron (MLP). The MLP was used to model energy and radiation fluxes and was compared to a reference ULSM (Town Energy Balance model). The MLP was found to be more accurate and more stable, especially in online mode. It can be concluded from these studies that the advantage of ML models lies in their lower computational cost once trained. Nevertheless, an emulated ML model can never exceed the accuracy of a numerical model, as it is trained on its model results.

This study proposes a novel and fast computational ML approach to model outdoor human thermal comfort at 1 x 1 m resolution in complex urban areas, hereafter called the Human Thermal Comfort Neural Network (HTC-NN). The HTC-NN can be used to downscale numerical weather prediction models or reanalysis data, considering urban geometry and function, and predict outdoor human thermal comfort at high resolution at a limited computational cost. The HTC-NN consists of four submodels: two neighborhood scale MLP models for modeling T_a and RH; a building-resolved U-Net model for modeling T_{mrt}; and a building-resolved statistical wind field. We use the MLPs to emulate a surface energy balance model to model T_a and RH at the neighborhood scale (500 x 500 m) at 2.0 m a.g.l. and a U-Net to emulate a T_{mrt} model at building-resolved scale (1 x 1 m) at 1.1 m a.g.l. (Briegel et al., 2023). In addition, Large-Eddy Simulations (LES) are emulated by a random forest (RF) model to compute statistical wind fields at 1.0 m a.g.l. in the urban canopy layer with a 1 x 1 m resolution for the four cardinal wind directions in relation to the forcing data. The wind fields are calculated from the x, y, and z wind components. The single-layer model Surface Urban Energy and Water balance Scheme (SUEWS) is used as surface energy model for the neighborhood scale modeling of T_a and RH (Järvi et al., 2011; Ward et al., 2016; Sun and Grimmond, 2019). SOLWEIG is the building-resolved T_{mrt} model (Lindberg and Grimmond, 2019). The U-Net that emulates SOLWEIG has already been developed and validated (Briegel et al., 2023). LES are used to obtain the scaling matrices and maps of mean wind speed in relation to forcing wind speed based on the LES of Albertson and Parlange (1999a, 1999b).

The objectives of this study are (i) to develop and validate ML models emulating numerical urban climate models to predict T_a, RH, and U at different scales, (ii) to link these models to predict and validate spatially distributed UTCI at 1 x 1 m resolution, and (iii) to map UTCI at high resolution in a case study of an entire urbanized area (Freiburg, Germany) over many years to derive a climatology of intra-urban variability of outdoor thermal climate.

2 Data

2.1 Study area

The study area covers the urbanized area of Freiburg im Breisgau, Germany, and is partially identical to the study area reported in Briegel et al. (2023). According to the Local Climate Zone (LCZ) classification, the urbanized area of Freiburg can be mainly classified into LCZs 5 (open midrise), 6 (open lowrise), and 8 (large lowrise), while some parts of the city center are LCZs 2 (compact midrise) and 3 (compact lowrise) (Stewart and Oke, 2012; Demuzere et al., 2022). Due to different training requirements, the training and testing model domains for the individual submodels differ, as well as the HTC-NN prediction domain. The MLP submodel training domains have an extent of 15 x 15 km and a grid size of 500 m, resulting in 436 grid cells. The training domain of the MLPs is larger than the actual HTC-NN prediction domain. The RF U submodel training domain covers 15 areas of varying grid size ranging from 122 x 122 m to 500 x 500 m. The final HTC-NN model prediction domain has an extent of 10 x 7 km and a grid size of 1 x 1 m. An overview of the training and prediction domains, individual training and test areas, and the urban sensor network is given in Fig. 1.

2.2 Spatial and forcing data

Spatial and forcing data are largely similar for the numerical and their corresponding ML submodels, however, the MLPs use only a subset of all SUEWS input data (see Sect. 2.4), and the RF uses additional spatial predictors. The derivation of all spatial data other than anthropogenic dynamics is described in Briegel et al. (2023). A detailed overview of the spatial input and forcing data required for each submodel and their derivation is given in Table 2.

SUEWS does not require explicit building-resolving data but local-scale averaged spatial data. Information on urban morphology, land cover classes, and anthropogenic dynamics are required for SUEWS. Population density for each grid cell is derived from demographic data by city district (City of Freiburg im Breisgau - Bevölkerung, 2022). The remaining spatial input data, such as emissivity and albedo, are left at default (Sun et al., 2021; Sun and Grimmond, 2019).

LES and SOLWEIG require three-dimensional building geometry and tree characteristics derived from digital surface models (DSMs), digital elevation models (DEMs), and building outline data (Briegel et al., 2023), hereafter referred to as DSMb and DSMv.

Measurements from the urban weather station of the University of Freiburg are used as meteorological forcing data. The weather station is located on a rooftop (~ 55 m a.g.l.; 48.0011, 7.8486) close to the city center (Fig. 1). A detailed description of the urban weather station can be found in Briegel et al. (2023). The following variables are used as forcing data for SUEWS, the corresponding MLPs, and the U-Net: T_a, RH, atmospheric pressure, downwelling shortwave radiation, downwelling longwave radiation, precipitation, U and wind direction. LES and the RF model require only standard forcing related to an initial shear velocity of 1 m s^{-1}.

2.3 Validation data

Validation data for the MLPs and HTC-NN are derived from an urban weather sensor network in Freiburg (Fig. 1), which was installed in the summer of 2022. The sensor network covers the entire urbanized area as well as some rural areas, adjacent valleys, and hills to account for local weather phenomena such as mountain-valley wind systems or altitude effects which are not resolved by the ML model.

The stations of the sensor network can be divided into Tier I stations, or biometeorological stations (7 stations), and Tier II stations (30 stations). Tier I stations measure T_a, precipitation, RH, wind speed and direction, global radiation, and black globe temperature (Feigel et al., 2023), while Tier II stations measure only T_a, precipitation, and RH (Plein et al., 2023). Tier I stations are equipped with full weather sensors (ClimaVUE50, Campbell Scientific Inc., Logan, UT, USA) and black globe temperature sensors (model BLACKGLOBE-L, Campbell Scientific Inc., Logan, UT, USA). All sensors are mounted at 3.5 m a.g.l. on streetlights or custom poles, and the measuring interval is one minute. Sensor network data in this study is aggregated to hourly values.

Table 1: Overview of the abstract input data, the spatial data and the method used to derive the abstract input data. Most calculations are performed by the Urban Multi-scale Environmental Predictor (UMEP - Lindberg et al. (2018)). Spatial data contains Land

Cover Classes (LCC), Digital Elevation Model (DEM), Digital Surface Model (DSM). Suffix "b" indicates building data and "v" vegetation data.

| Abstract spatial input	Spatial data	Method
Surface cover fractions	LCC map	UMEP Land Cover Fraction
Zero displacement height (b / v)	DEM, DSMb / DSMv	UMEP Morphometric Calculator: (Kanda et al., 2007)
Roughness length (b / v)	DEM, DSMb / DSMv	UMEP Morphometric Calculator: (Kanda et al., 2007)
Mean height (b / v)	DEM, DSMb / DSMv	UMEP Morphometric Calculator: (Kanda et al., 2007)
Frontal area index (b / v)	DEM, DSMb / DSMv	UMEP Morphometric Calculator: (Kanda et al., 2007)
Altitude	DEM	Mean height of grid cell
Population density	Population data by district	Weighted mean of different districts in specific grid cell

Besides the sensor network, SOLWEIG is run for small subsets (50 x 50 m) around the Tier I stations to derive UTCI and T_{mrt} data for the specific locations of the Tier I stations (POI function - Lindberg and Grimmond (2019)). This allows a more detailed evaluation and a better attribution of the HTC-NN results.

2.4   Study period

The study period is aligned with the sensor network data which are collected from June 2022 onwards. The entire model period is from 2018 to 2022, with 2018 used as a spin-up year for SUEWS. The MLPs are trained with data from 2019 to 2021 and tested for 2022 (January–December). Model validation of SUEWS and the MLPs are performed with measurement data from June to December 2022, while the HTC-NN is validated with data from August to December 2022. During the study period, the mean annual T_a is 13.00°C, the mean summer (June-August) T_a is 21.32°C, and the mean maximum daily summer T_a is 26.27°C. The number of hot days (maximum $T_a >= 30$°C) of the consecutive years from 2019 to 2022 are 26, 20, 9, and 37, respectively.

3   Modeling approach

The development of the HTC-NN requires four steps (Fig. 2). The first step is to generate initial spatial and meteorological data from various sources, which are listed in Table 2. In the second step, the so-called 'ground truth' data (T_a, RH, T_{mrt}, and U) for the four HTC-NN submodels (two MLPs, U-Net, and RF) are calculated using numerical models (SUEWS, SOLWEIG,

5
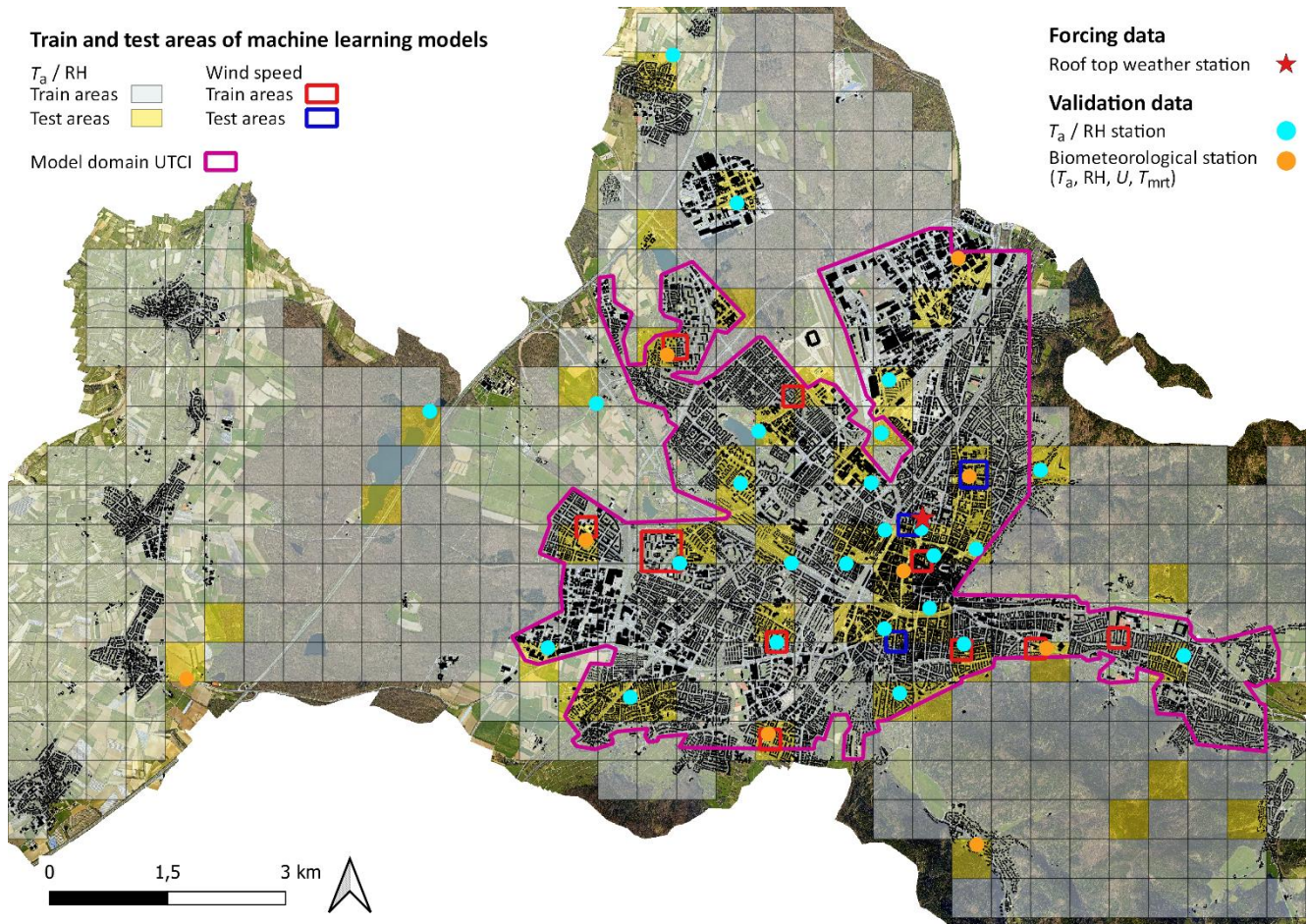
Figure 1: Model domain of the City of Freiburg, Germany. The red star shows the location of the weather station used for forcing data on a rooftop. Orange and turquoise points show the locations of the urban sensor network used for model evaluation. Gray grid cells show the training areas of the T_a and RH submodels, while yellow grids show the test areas. Red and blue squares show the training and test areas of the U sub model, respectively. The pink border shows the prediction area of UTCI. Orthophoto in the background based on data from the City of Freiburg, www.freiburg.de.

and LES). Training and evaluation of the HTC-NN submodels are done in the third step, while the fourth step is to link these submodels by calculating UTCI. As the U-Net has already been trained and validated, only the development and the requirements of the MLPs and the RF (spatial and temporal data, SUEWS, and LES) are explained.

3.1 Numerical Modelling

3.1.1 Local scale T_a and RH modeling (SUEWS)

SUEWS (version 2020a) is used to model T_a and RH at 2.0 m a.g.l. for 436 grid cells with a resolution of 500 x 500 m (Järvi et al., 2011; Ward et al., 2016). SUEWS has been validated in different cities under different climatic conditions (Ward et al.,

2016; Järvi et al., 2011; Ao et al., 2018). Besides the following parameters, SUEWS is run in default mode: net radiation method, maximum and minimum porosity, roughness length of momentum method. Netradiation method is set to 1, as downwelling longwave radiation data are available. Maximum and minimum porosity of deciduous trees is set to 0.6 and 0.2, respectively (Ward et al., 2013). Roughness length and zero displacement height are calculated according to Kanda et al. (2007) and are provided to SUEWS, so the roughness length method is set to 1. As mentioned, SUEWS is run for 2018–2022, while 2018 is excluded from subsequent modeling as it serves as a spin-up year.

Table 2: Overview of required spatial and forcing data for the numerical and machine learning models. Note: SUEWS and MLP use abstract spatial data (see Table 1) and the RF model uses additional spatial predictors derived from DEM, DSMb, and DSMv which are not listed here.

Data	SUEWS / MLP (500 x 500 m)	SOLWEIG / U-Net (1 x 1 m)	LES / RF (1 x 1 m)
LCC map	x	x	-
DEM	x	x	x
DSMb	x	x	x
DSMv	x	x	x
Sky view factor	-	x	-
Wall height and aspect	-	x	-
Soil characteristics	x	-	-
U	x	x	x
T_a, RH, atmospheric pressure, downwelling shortwave radiation, downwelling longwave radiation, precipitation, wind direction	x	x	-

3.1.2   Micro-scale U modeling (LES)

LES are used to obtain micro-scale (1 x 1 m) wind fields for 15 areas ranging in size from 122 x 122 m to 500 x 500 m. The LES model code and set-up are described in Giometto et al. (2016) and (2017). LES needs DSMb and DSMv as spatial input data. Training and test areas are carefully selected to ensure that the variability of the urban environment is adequately represented. For each area, four LES are computed for pressure gradients that cause north, east, south, and west inflow directions. The simulations are set up with an identical standard forcing related to an initial shear velocity of 1 m s^{-1}. Each simulation contains 30 minutes (steady-state) wind field data in two-second steps, which are then time-averaged.
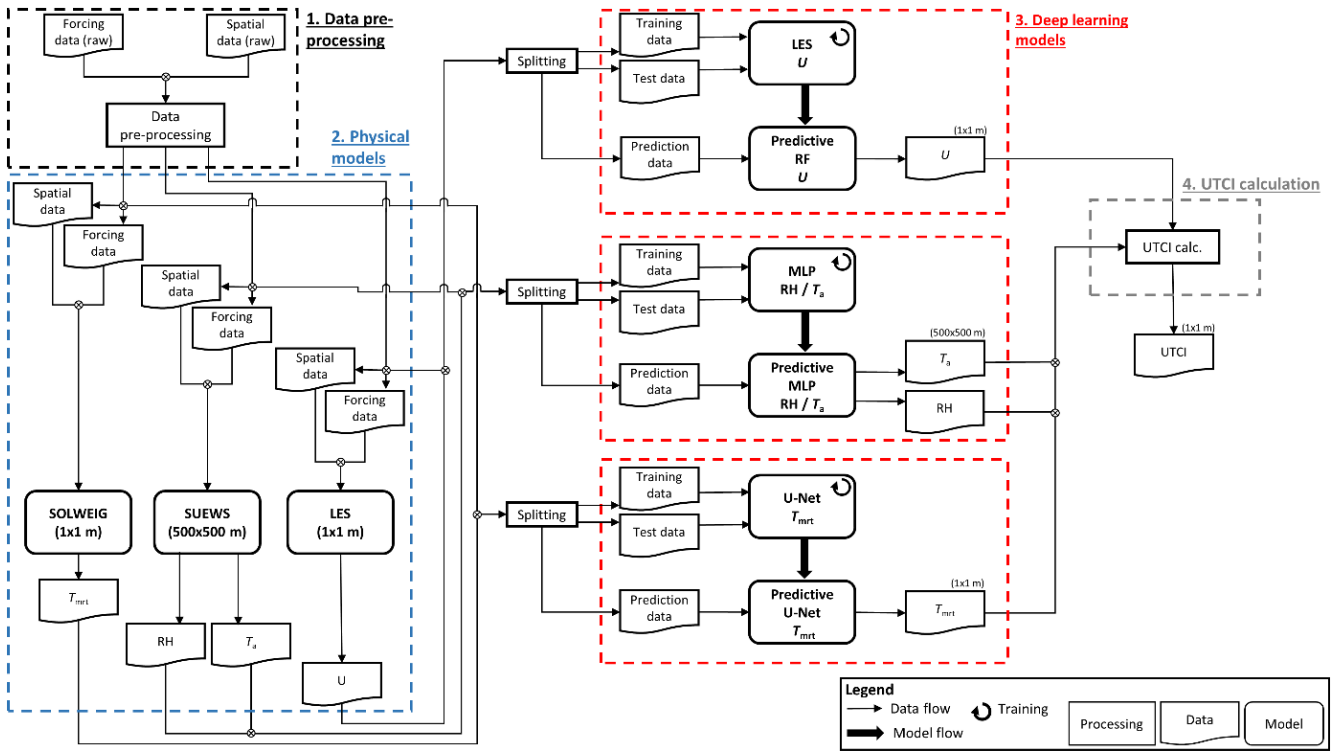
Figure 2: Workflow of the HTC-NN model development: 1. Data pre-processing; 2. Numerical modeling of 'ground truth'; 3. Training and evaluation of machine learning models; and 4. Calculation of UTCI.

3.2 Machine learning models

3.2.1 Multi-layer perceptron model development

175 The MLPs are designed using Python and the PyTorch library. To determine the best model architecture, a model-based Bayesian hyperparameter optimization is performed. For this purpose, the Optuna software framework (Akiba et al., 2019) is used in combination with PyTorch. Training and prediction are conducted on an NVIDIA GeForce RTX 3080.

T_a and RH are each modeled with their own deep learning model. Different to the T_{mrt} deep learning model, which is based on a convolutional neural network, T_a and RH models are built from fully-connected feedforward artificial neural networks,

180 known as MLP. This is because the different SUEWS grid cells are not connected and have no spatial relationship. MLPs consist of three different types of layers: input layer, hidden layers, and output layer. Bayesian optimization determines the number of hidden layers and neurons, the presence/absence of a dropout layer, and the learning rate that leads to the highest model accuracy. In total, 30 hyperparameter combinations (trials) are tested for each model. For this purpose, the training dataset is further divided into training and evaluation datasets, and a 3-fold cross-validation is used to validate each

185 hyperparameter trial. The hyperparameter combination of each trial is defined by the Tree-structured Parzen Estimator (TPE) (Bergstra et al., 2013, 2011), which is a Bayesian hyperparameter optimization algorithm. The best architectures of both MLPs

have three hidden layers with 60, 49, and 42 neurons for the T_a MLP and 34, 53, and 54 neurons for the RH MLP. Learning rates are best at 0.0014 and 0.0011 for the T_a MLP and RH MLP, respectively. Drop-out layers do not improve the model accuracy and are not added to the final models. The remaining hyperparameters, such as activation function, optimizer, and initial weight distribution, are taken from the literature (Table 3 and Briegel et al. (2020)). As an evaluation metric (loss function), mean squared error (MSE) is used. For further comparisons, root mean square error (RMSE) as well as mean bias error (MBE) are used.

3.2.2   Random Forest model development (U)

The RF modeling is conducted in MATLAB (The Mathworks, Natick, MA, USA, version 2021b) using a regression-bagging approach with 50 learning cycles. In addition to spatial building and tree data, spatial features (derived from DSMb) are predictors. These predictors are indices pertaining to the street length and width, frontal area index relative to the flow direction, horizontal Euclidean distance, and up- and downwind distances to buildings. For each of the 12 training areas and the three components of the wind field, individual models are trained (u: streamwise, v: normal-to-streamwise horizontal, w: vertical), resulting in 36 models. This model ensemble makes predictions for each component, which are then assembled as the Euclidean norm of u, v, w to obtain the final wind field. The RF models can theoretically compute wind fields for any wind direction over the city, requiring only an adjustment of the predictors. However, for the purpose of calculating outdoor thermal comfort, citywide wind fields are computed only for the north (315–45 °), east (45–135 °), south (135–225 °), and west (225–315 °) inflow directions, each covering a 90° angle. The LES wind fields are scale-independent, allowing a linear rescaling of the RF

Table 3: Hyper-parameters of both MLP models.

Hyper-parameter	Value
Activation function	ReLU
Optimizer	Adam
Initial weight distributions	He uniform
Loss function	MSE
Epochs	20
Batch size	62
Hidden layers	T_a MLP: 3
	RH MLP: 3
Neurons	T_a MLP: 60, 49, 42
	RH MLP: 34, 53, 54
Learning rate	T_a MLP: 0.0014
	RH MLP: 0.0011

9

model output. Thus, a time series of city-wide wind fields can be computed by hourly wind speed data from the urban weather station.

3.3   Thermal Indices

In the final model step, the results of the above-mentioned submodels are combined into a thermal comfort index with a spatial resolution of 1 x 1 m. However, not all indices are appropriate for human thermal comfort (Staiger et al., 2019). Therefore, only UTCI (Błażejczyk et al., 2013), which is widely used in urban climate science and planning, is considered in this study. The reference conditions of UTCI correspond to an individual walking outdoors with T_{mrt} equal to T_a, no wind and RH at 50 %. The UTCI values can be categorized based on thermal stress, e.g., UTCI values ranging from 32–36°C are assigned to strong heat stress. The different UTCI stress categories and the corresponding UTCI ranges are listed in Table A1.

4 Result

This section presents the evaluation of the three submodels of the HTC-NN (Sect. 4.1), of the HTC-NN itself (Sect. 4.2) and the high-resolution UCTI mapping of the HTC-NN (Sect. 4.3).

### 4.1 Evaluation of T_a, RH, and U submodels

220 The accuracies of SUEWS, SOLWEIG, the MLPs, U-Net, and RF relative to the sensor network are shown in Fig. 3 and Table 4. A spatial evaluation of machine learning models to numerical models on their specific test areas is given in Table B1.

With an RMSE of 1.50 K and 1.51 K respectively, the SUEWS and T_a MLP models have similar performance compared to the sensor network data. Both models have a lower RMSE of about 0.3 K compared to the forcing data. The errors of SUEWS and the MLP across the different stations are similar (Fig. 3a), with a higher variability during the night than the during day.

225 Figure 4b shows a moving average of RMSE over time of the final HTC-NN model and of the T_a and RH MLPs (Fig. 4c and d). It can be seen that the RMSE of the T_a MLP shows two peaks in late October and mid-December, while the RMSE fluctuates around 1 K during the remaining time. Overall, the T_a MLP shows good accuracy to the SUEWS model (R^2 of 0.997 and RMSE of 0.5 K). Similar observations can be made for the RH MLP. RMSE for SUEWS and the MLP model are 7.79 % and 8.14 %, respectively, while the RMSE of the forcing data is 9.08 %. Similar to the T_a models, the RMSE of the RH MLP has

230 two strong peaks in late October and mid-December with RMSE values up to 15 %, while RMSE fluctuates

Table 4: MAE, MBE, RMSE, and coefficient of determination (R^2) of numerical / machine learning models and sensor network data for T_a, RH, U, T_{mrt}, and UTCI. Additionally, errors between forcing data and sensor network data are added for T_a, RH, and U (as baseline). Note, T_a and RH are validated on Tier I and Tier II stations, while U, T_{mrt}, and UTCI are only validated on Tier I stations.

| Variable	Model / Data	MAE	MBE	RMSE	R^2
T_a	SUEWS	1.07 K	-0.15 K	1.50 K	0.97
	MLP	1.08 K	-0.17 K	1.51 K	0.97
	Forcing	1.32 K	-0.12 K	1.83 K	0.95
RH	SUEWS	5.80 %	-0.44 %	7.97 %	0.88
	MLP	5.99 %	-0.33 %	8.14 %	0.87
	Forcing	6.54 %	-1.52 %	9.08 %	0.84
U	RF	0.52 m s^{-1}	0.24 m s^{-1}	0.73 m s^{-1}	0.28
	Forcing	2.27 m s^{-1}	-2.24 m s^{-1}	2.85 m s^{-1}	0.35
T_{mrt}	U-Net	4.25 K	-1.53 K	6.18 K	0.84
	SOLWEIG	3.83 K	-0.97 K	5.86 K	0.86
UTCI	HTC-NN	2.27 K	1.34 K	3.00 K	0.92
	SOLWEIG	2.48 K	-0.66 K	3.29 K	0.90
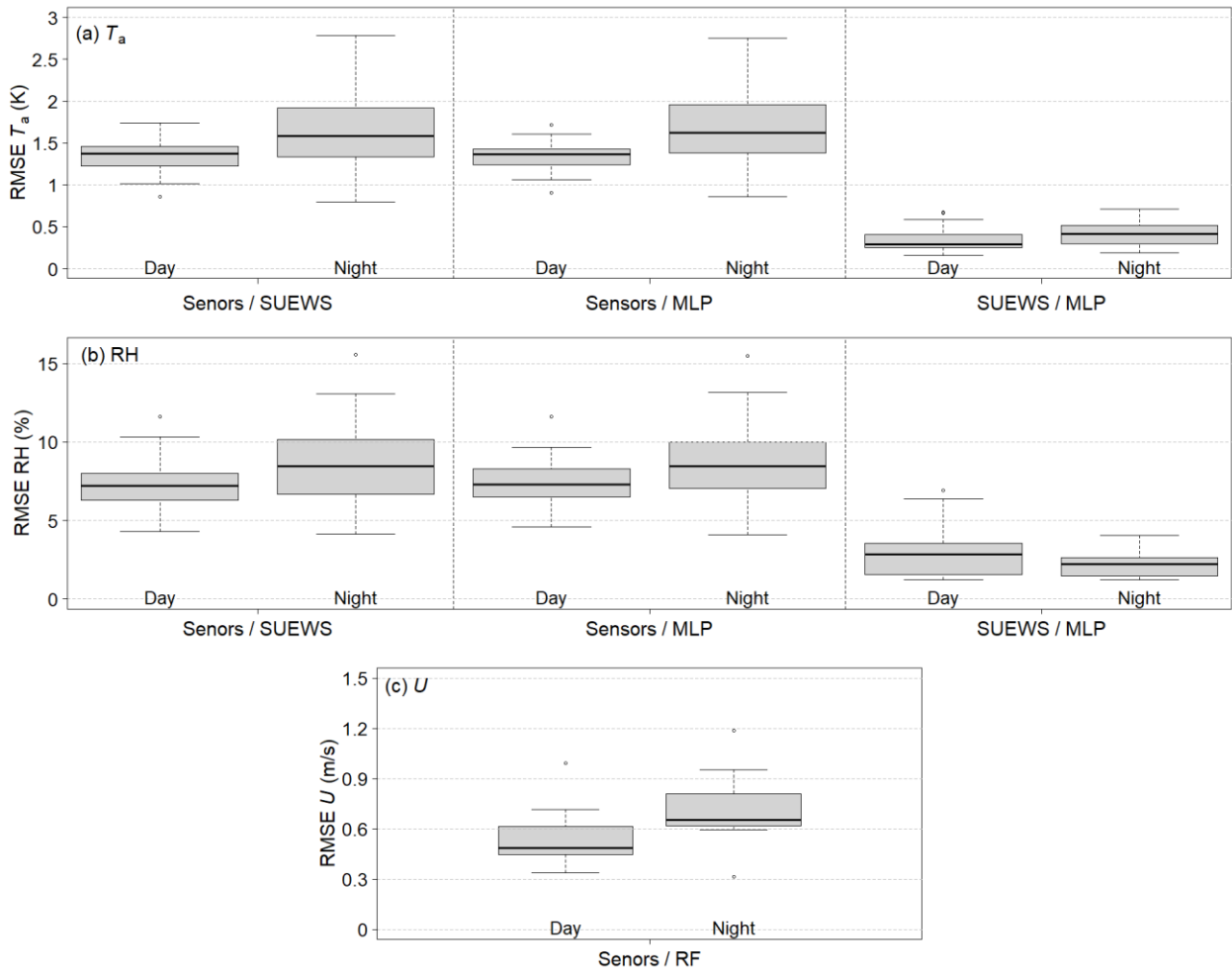
Figure 3: Boxplots of RMSE of SUEWS and MLPs compared to sensor network data and between SUEWS and the MLPs for T_a (a) and RH (b) respectively and their differences between day and night. In (c) RMSE of U between sensor network data and RF is shown. Boxplots show dispersion of different sensor network stations.

235

around 5 to 6 % the remaining time. R^2 of SUEWS and the RH MLP are lower at 0.88 and 0.87 compared to 0.97 of the T_a models. Nevertheless, the overall R^2 between SUEWS and the RH MLP is high at 0.98, and with an overall RMSE of 3.28 %, the accuracy of the RH MLP is considered satisfactory. The T_{mrt} U-Net has a slightly lower accuracy than SOLWEIG (RMSE of 6.18 K to 5.86 K; R^2 of 0.84 to 0.86). A detailed evaluation can be found in Briegel et al. (2023). The RF has an R^2 of 0.28

240 and an RMSE of 0.74 m s^{-1} in relation to sensor network data and shows a large improvement to forcing data (RMSE of 2.85 m s^{-1}). However, R^2 of RF U is lower than R^2 of forcing data. The RF model has an overall accuracy of 0.76 m s^{-1} (RMSE) compared to the LES model (Table B1).
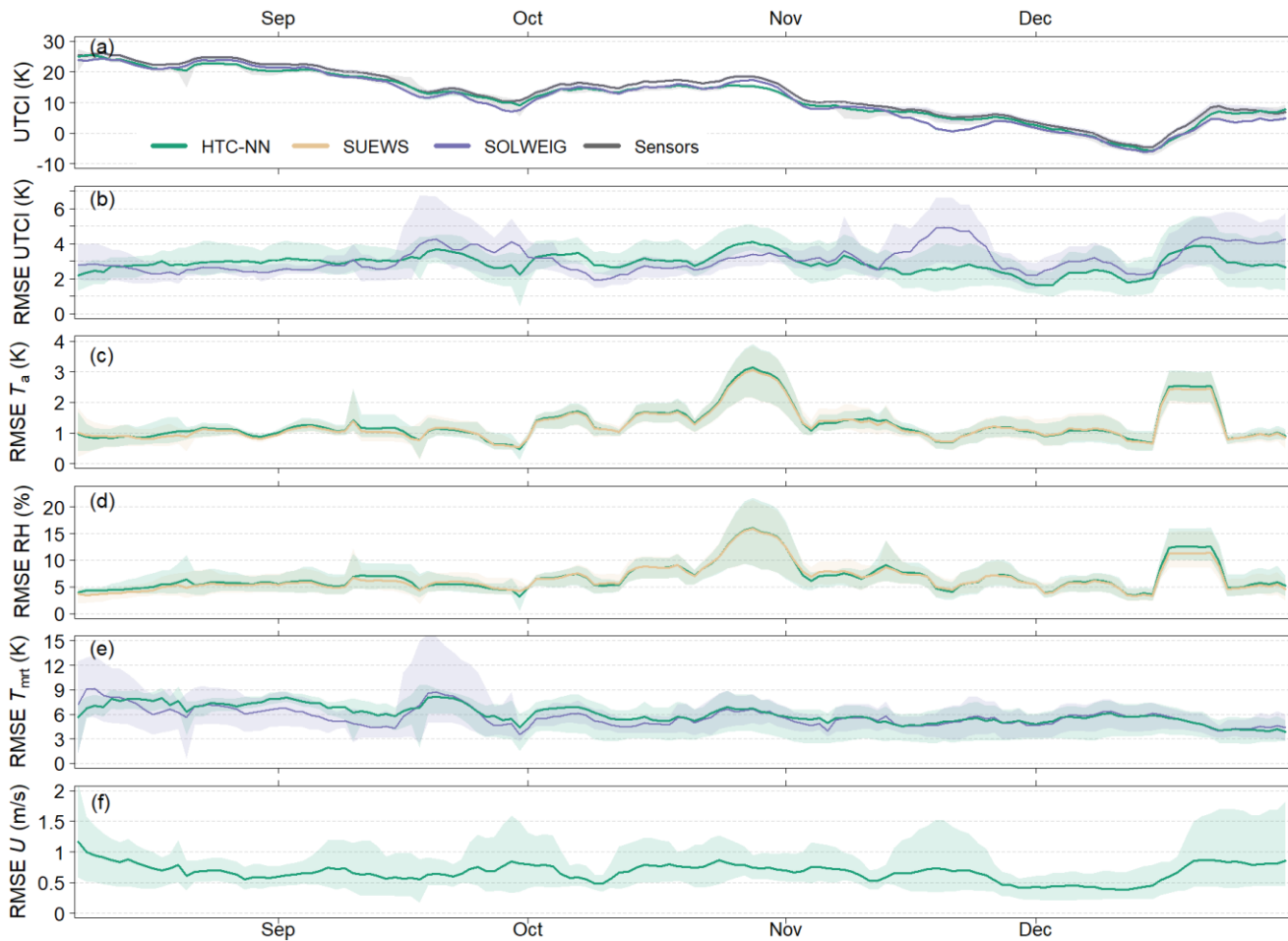
Figure 4: Moving average of UTCI (a) and RMSE of UTCI (b), T_a (c), RH (d), T_{mrt} (e), and U (f) from August to December 2022. The window size of the moving average is seven days. Time series starts with the installation of the first Tier I stations in August 2022. In (c) and (d) RMSE of SUEWS predictions are added as comparison (orange lines). In (b) and (e) RMSE of SOLWEIG predictions are added (violet lines). For U no numerical model results exists. Shaded areas represent 95 % confidence interval.

4.2 HTC-NN evaluation

245   The HTC-NN has an RMSE of about 3.00 K and an R^2 of 0.92 (in relation to sensor network), while SOLWEIG has an RMSE of 3.29 K and R^2 of 0.90 (Table 4). Moving averages of UTCI and RMSE (window size of 7 days) over time (August–December 2022) and diurnal error patterns are shown in Fig. 4 and 5. The annual RMSE of UTCI ranges between 2 and 4 K, with the highest RMSE values in mid-September, late October, and mid-December. T_a and RH show a similar pattern in October and December, while T_{mrt} has its largest errors in August and September, ranging from 6 to 8 K. The RMSE of U varies between

250   0.5 and 1 m s^{-1} with the highest values in August, late October, and late December. The elevated RMSE values of UTCI in mid-September coincide with the RMSE peak of T_{mrt} in mid-September, while the RMSE peaks of UTCI in late

13
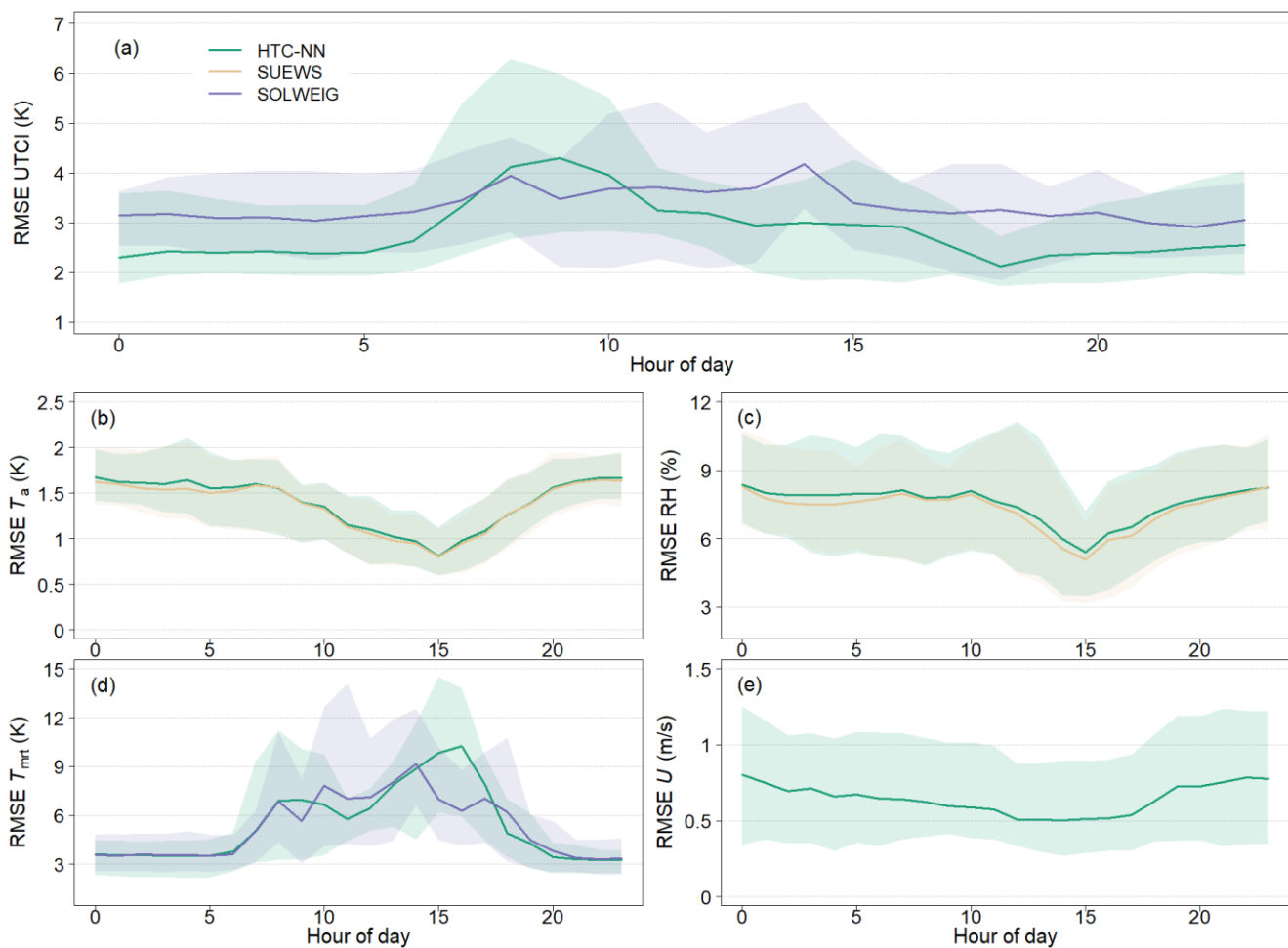
Figure 5: Normalized diurnal RMSE of UTCI (a), T_a (b), RH (c), T_{mrt} (d) and U (e). In (b) and (c) RMSE of SUEWS predictions are added as comparison (orange lines). In (a) and (d) RMSE of SOLWEIG predictions are added (violet lines). For U no numerical model results exists. Shaded areas represent 95 % confidence interval.

October and December match the high RMSE values for T_a and RH in those periods. In Fig. 4b and e, results from SOLWEIG are added (violet lines). SOLWEIG is more accurate than the HTC-NN in modeling UTCI / T_{mrt} in summer, while the HTC-NN is more accurate in autumn and winter. Diurnal patterns of UTCI accuracy are shown in Fig. 5a. Model errors are lower during the night than during the day, with the highest errors during the morning. The U-Net T_{mrt} error is also lowest during the night but has its highest errors in early evening (see also Briegel et al. (2023)). The T_a and RH MLPs have similar error patterns and are lowest in the late afternoon. The diurnal pattern of the U RF model accuracy shows higher accuracy during the day than at night.

Figure 6 shows sensor network and modeled UTCI values (HTC-NN and SOLWEIG) at five stations during an exemplary heat wave event in August 2022. Daytime UTCI is overestimated by the HTC-NN to varying degrees (between 1 and 5 °C), while nighttime predictions are in line with measurements. In Fig. 6c, the HTC-NN underestimated UTCI in the afternoon,

14
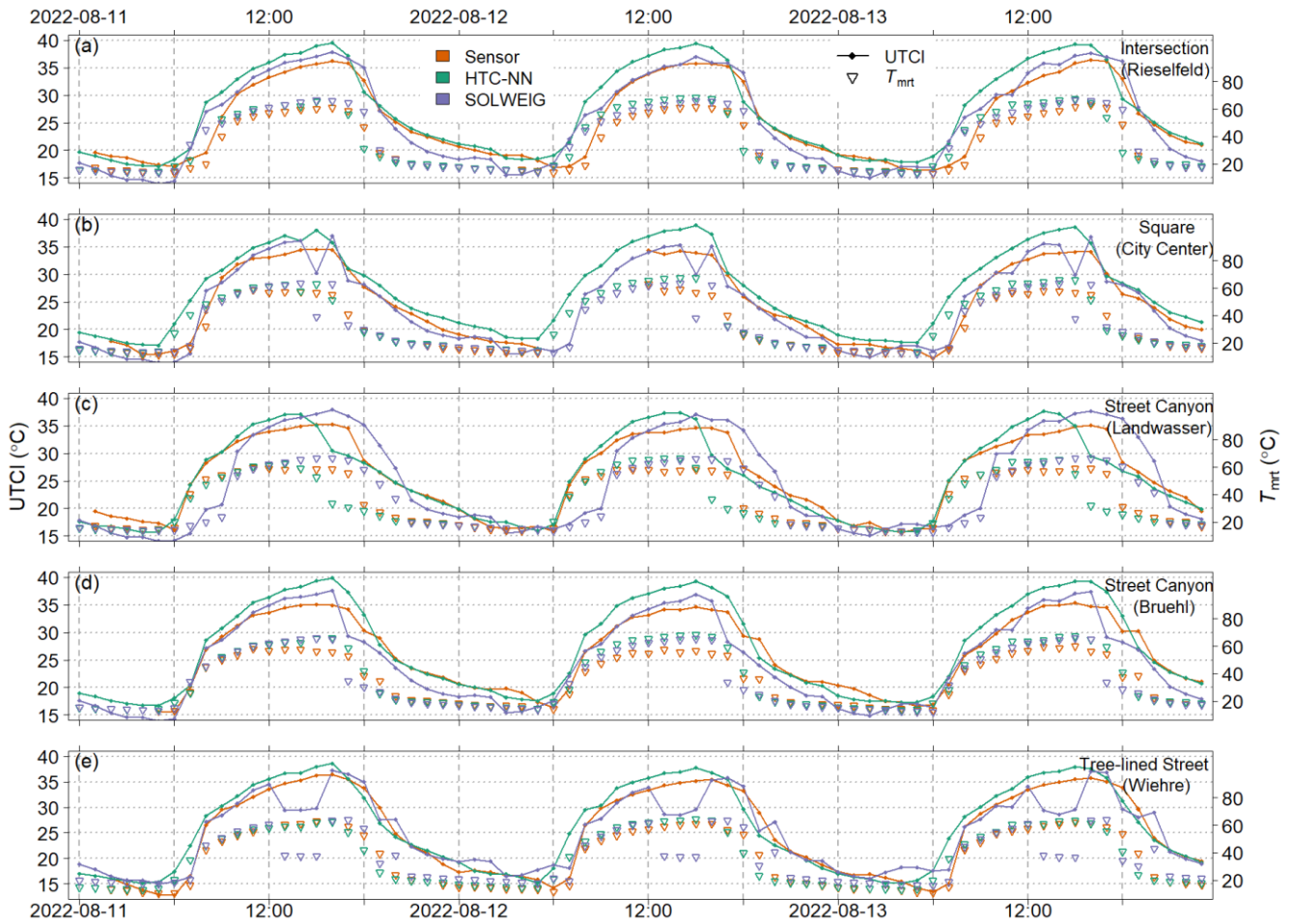
Figure 6: Modeled and measured UTCI and T_{mrt} at five Tier I stations (a)–(e) during a heatwave event in August 2022.

which is in line with underestimated T_{mrt}. SOLWEIG, on contrary, underestimates UTCI in the morning and overestimates it
in the afternoon and evening. Figure 6e shows significantly lower UTCI and T_{mrt} values of SOLWEIG during the afternoon,
while the HTC-NN does not show this pattern.

In Fig. 7, distributions of both sensors and models and their differences are shown. The HTC-NN has a higher share of values
between 15 and 25 °C, whereas shares of UTCI greater 25 °C are equal. SOLWEIG, on the contrary, has lower shares for
almost all the bins below 2.5 °C and higher shares between 5 and 12.5 °C.
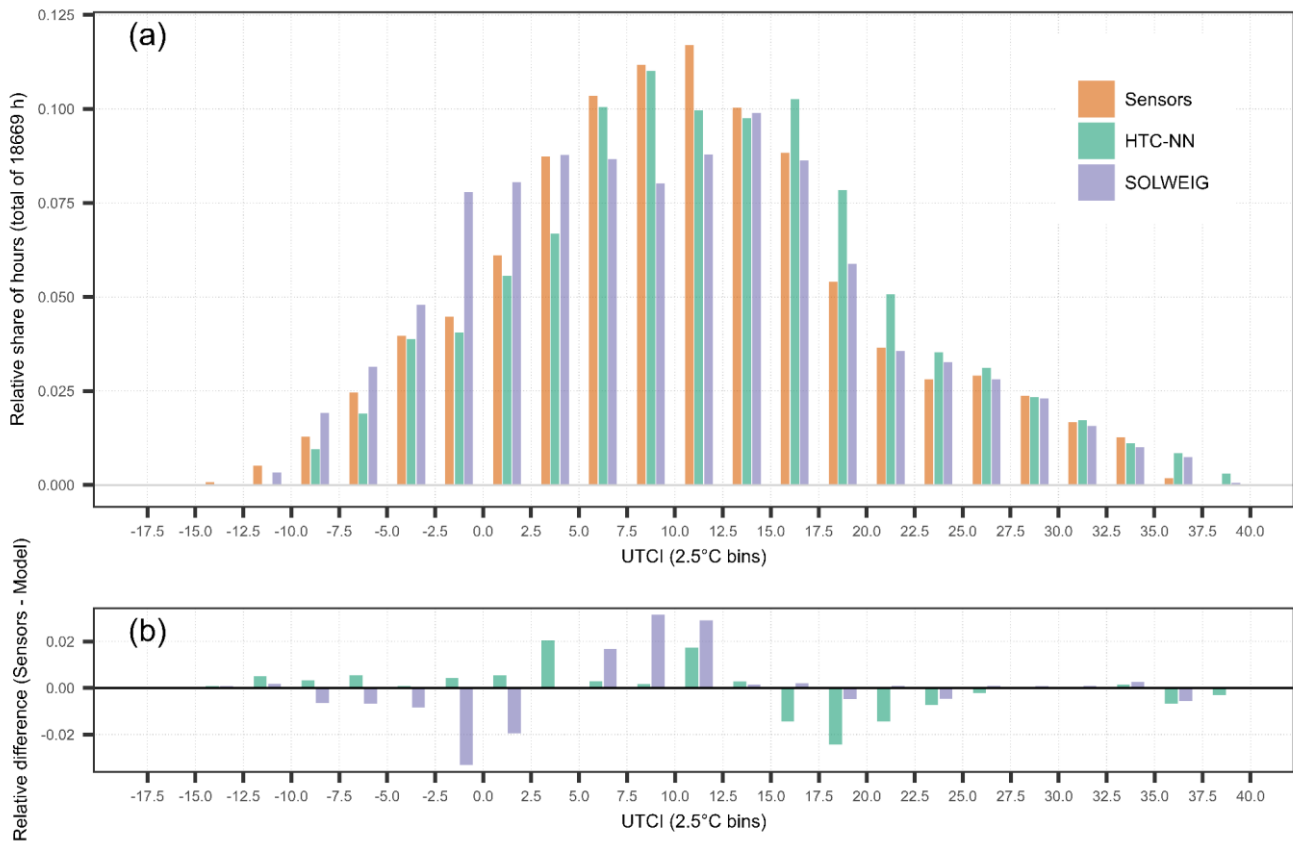
Figure 7: UTCI distributions from sensors, HTC-NN, and SOLWEIG (a) and the differences between models and sensor network (b). Distributions are shown for bins with a size of 2.5 °C and their relative share on total hours (n = 18669 h).

4.3 High-resolution UTCI mapping

275 This section presents an application of the developed HTC-NN for mapping and downscaling simulated UTCI at 1 m resolution for Freiburg. Calculating hourly UTCI over four years (35064 timesteps) and 42.5 million pixels using an Intel Core i9 processor and NVIDIA GeForce RTX 3080 took only about 8 days. No individual runs are saved as this would exceed storage capacities, and predictions are stored as day- and nighttime sums of hours for each month within 1 °C UTCI bins.

Figure 8 shows the spatial distribution of cumulative daytime hours with strong, very strong, or extreme heat stress. This

280 figure shows heat hotspots related to daytime heat stress, as UTCI > 32 °C are rarely reached at night without solar radiation (Fig. 11b). Figure 10 shows the same map extent but with nighttime hours exceeding a UTCI value of 22 °C. Figures 9 and B3Figure B3 provide a more detailed view of four exemplary urban areas representing the LCZs 2, 5, and 8. These areas show not only the thermal comfort difference between different LCZs but also the intra-LCZ variability (Fig. 9c and 9d). The four areas

16
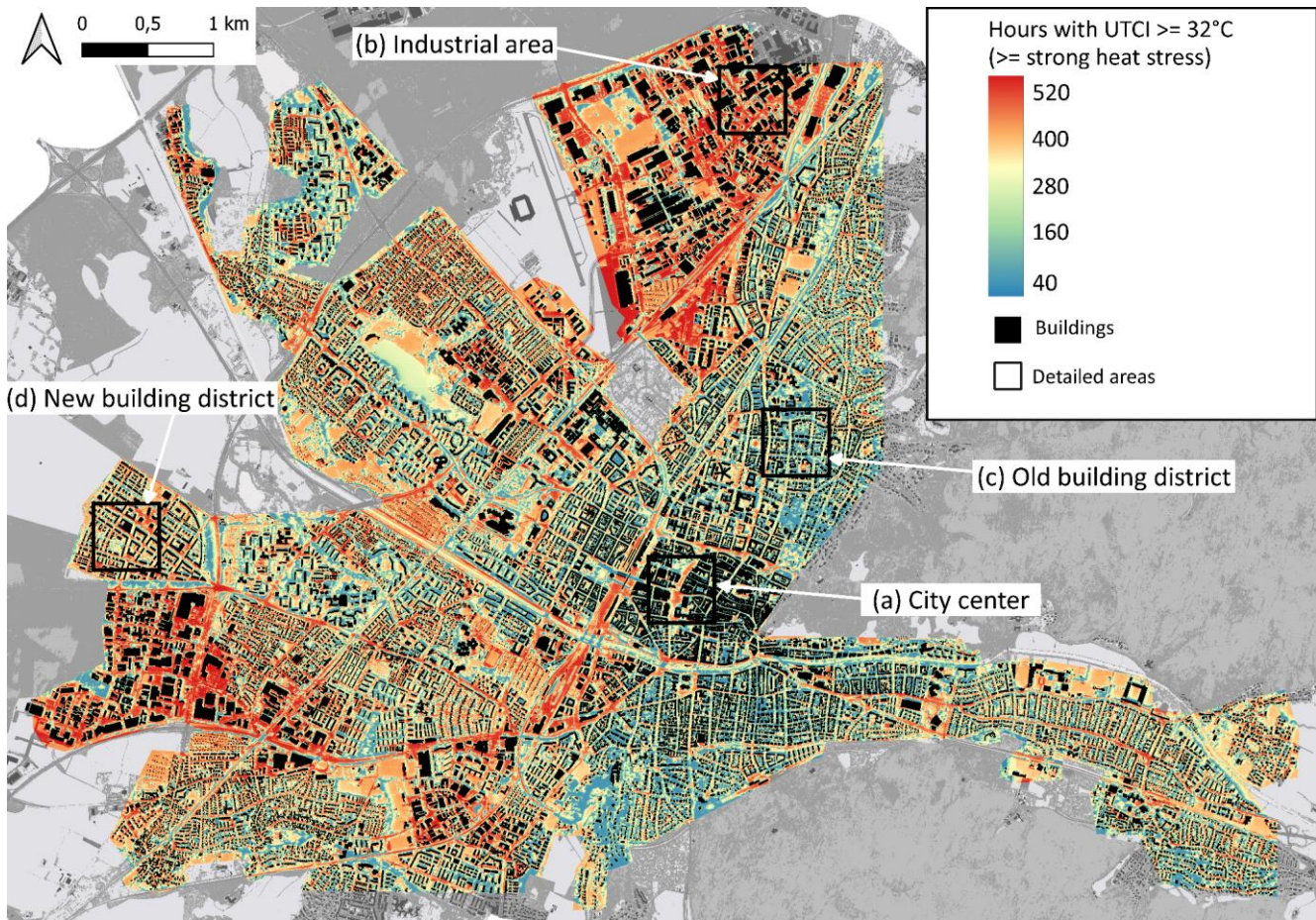
Figure 8: Map of Freiburg representing the average amount of daytime hours with a UTCI ≥ 32 °C per pixel (corresponds to strong and more heat stress). Note, coloring is in accordance with predicted UTCI quantiles. Spatial resolution of the map 1 x 1 m and time period for averages is 2019–2022. Rectangles (a)–(d) define areas shown in more detailed in Fig. 9, 11, and B3Figure B3.

285

represent the densely built-up city center with partly large trees but also large open and paved areas (a), an industrial area with mostly paved surfaces and only a few trees (b), an old building district with a large and old tree stock (c), and a relatively new district built 1995–2005 (d).

In Fig. 9, it is quickly apparent that there are large differences due to shading and tree canopy during the day and night.

290 Figure 9e and B3Figure B3e show the distributions of the spatially distributed amounts of hours in Fig. 9a–d. The distributions confirm the visual perspective. The industrial area (b) and the new building district (d) have a higher share of pixels, with more hours exceeding the threshold of 32 °C and fewer hours exceeding the threshold of 22 °C during the night. In turn, the densely built-up city center (a) and the old building district (c) have fewer hours ≥ 32 °C but more hours > 22 °C. Figure 11 shows the summertime UTCI distributions during the day- and nighttime for these four areas. The distributions underline the visual

295 assumption that tree-covered and less-paved areas are less affected by thermal stress during the daytime, but have higher UTCI values during nighttime. In addition, Fig. 11 shows the 80th percentiles of the different areas, ranging from 29 °C (old

17
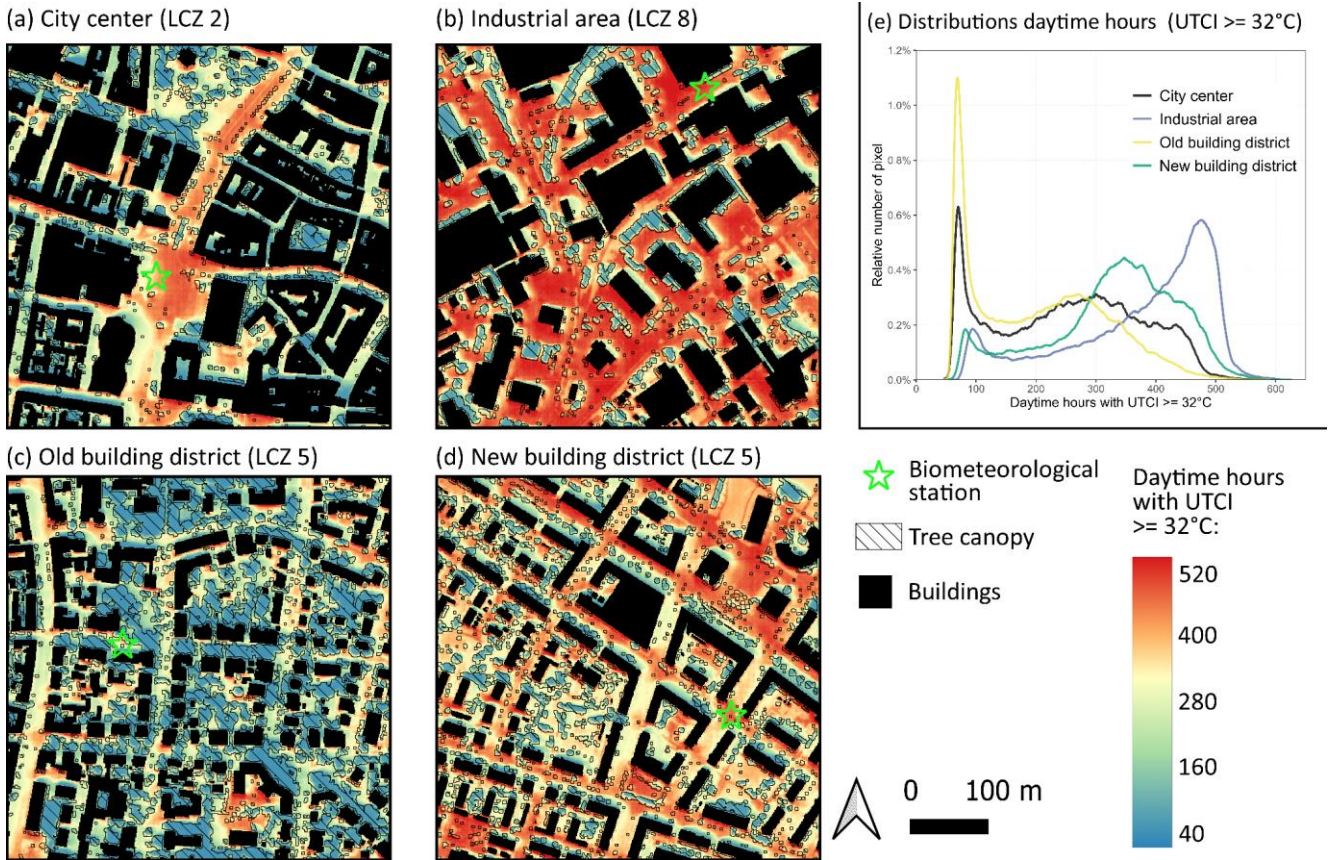
Figure 9: Four 500 x 500 m subsets of different urban neighborhoods of Freiburg. Average hours per year with a UTCI ≥ 32 °C are shown. (a) shows the city center of Freiburg, (b) an industrial area in the north of Freiburg, (c) a district with old buildings and dense / mature tree stock, and (d) a building district built after the year 1995. (e) shows the distributions of (a)–(e). Note, coloring is in accordance with predicted UTCI quantiles. Spatial resolution is 1 x 1 m and time period is 2019–2022.

building district) to 32 °C (industrial area) during the daytime and from 21 to 22 °C during nighttime. The 80th percentiles highlight the importance of shading to daytime outdoor thermal comfort. During the nighttime, however, the industrial area

300 and the new building district cool down more quickly, resulting in lower UTCI values and 80th percentiles (21 °C vs 22 °C). Although there is a cooling effect at night, the shading effect during the day is stronger than the reduction of nighttime cooling. These results indicate that outdoor thermal comfort controls have an inverted effect during the day and the night.

There are not only differences between the different LCZs but also within the same LCZ class. The old building and the new building district show strong differences in their day- and nighttime UTCI distributions. The difference between those two

305 areas is larger than between the city center and the old building district or between the industrial and new building districts. The areas do not differ much in their physical building characteristics and are both classified as LCZ 5. However, the old

18
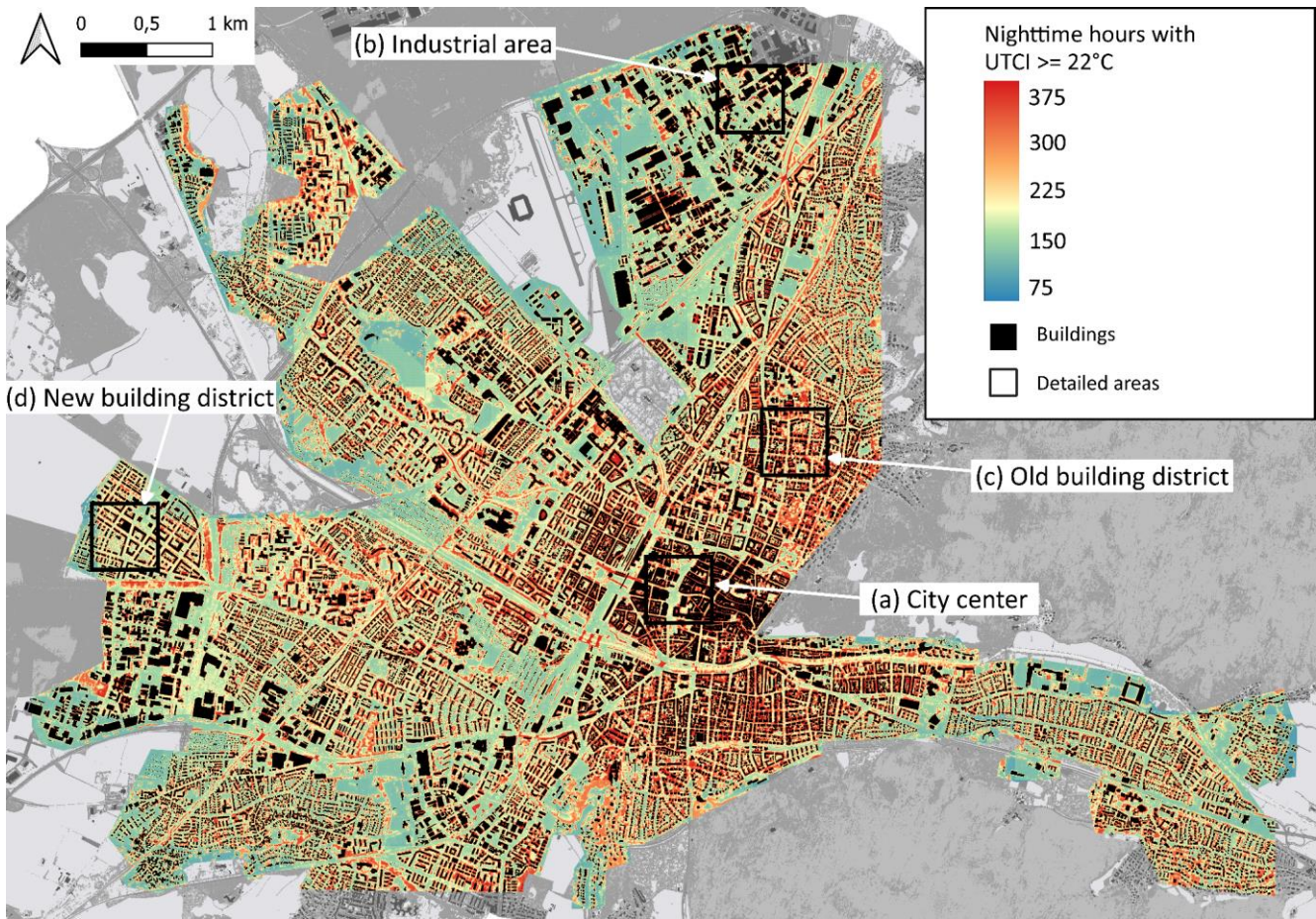
Figure 10: Map of Freiburg representing the average amount of nighttime hours with a UTCI ≥ 22 °C per pixel. Note, coloring is in accordance with predicted UTCI quantiles. Spatial resolution is 1 x 1 m and time period is 2019–2022. Rectangles (a)–(d) define areas shown in more detailed in Fig. 9, 11, and B3Figure B3.

building district has a tree cover fraction of 45 % and can be further defined as LCZ subclass 5_b, while the new building district has only 24 % tree cover and can therefore defined as LCZ 5_d.

310   In Fig. 10, sharp UTCI transitions are partially visible. These transitions occur between adjacent grid cells (500 x 500 m) of the T_a and RH MLPs (= original SUEWS grid cells). The transitions are more prominent in Fig. 10 than in Fig. 8, as T_{mrt} more influences diurnal UTCI than T_a. The transitions between grid cells also illustrate that T_a and RH are modeled individually for each grid cell without interaction with adjacent cells.
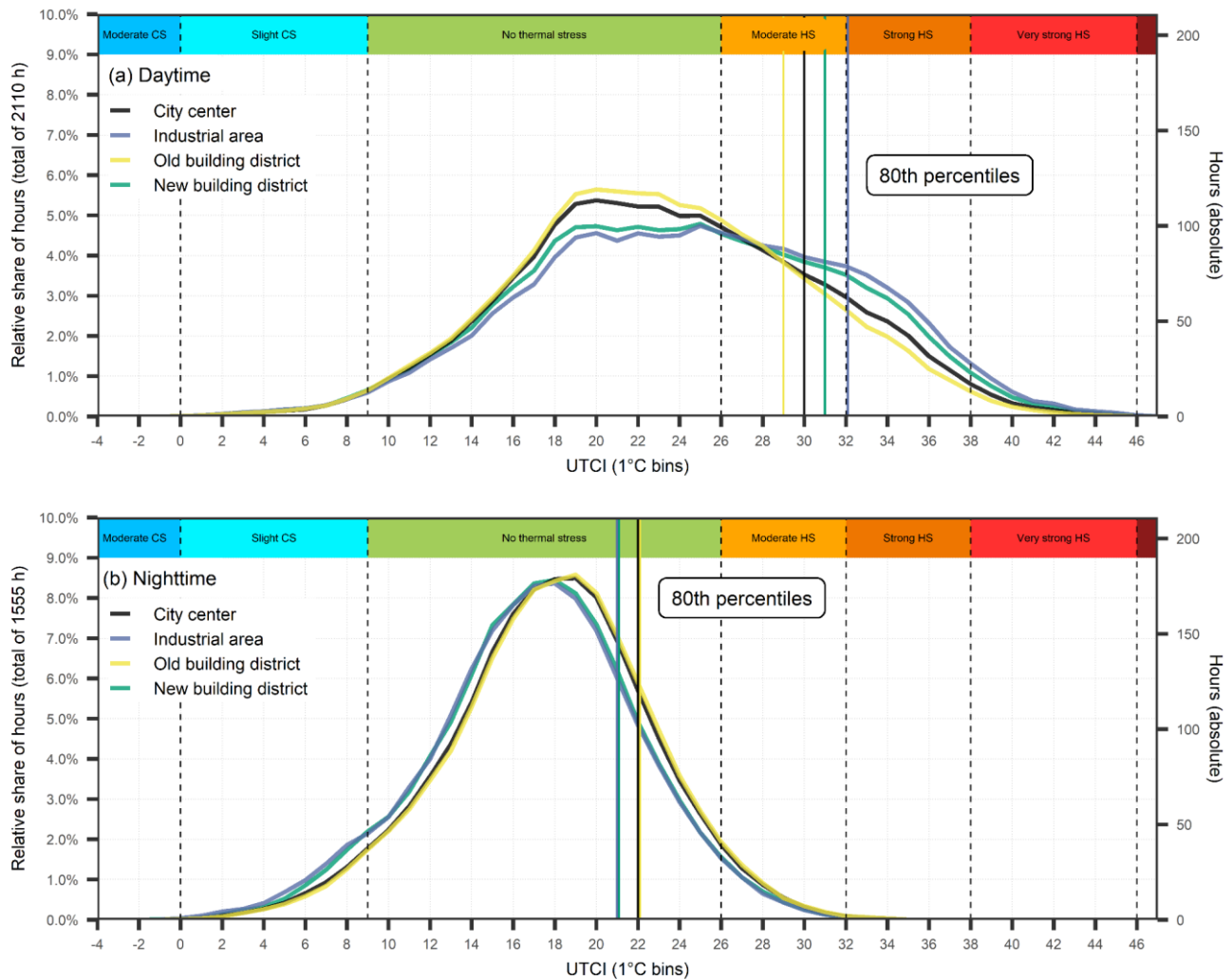
Figure 11: Summer time (May–September) distributions of UTCI of the four districts shown in Fig. 9 and B3Figure B3 at daytime (a) and during nighttime (b). On top of each figure the UTCI heat stress classes are presented (CS: cold stress; HS: heat stress). Vertical lines in the colors of the four areas represent the 80th percentiles. Left y-axis shows the relative and right y-axis the absolute number of hours. Note: x-axis has a bin size of 1 °C; total amount of hours varies between (a) and (b) because of day length in summer.

5   Discussion

5.1   Evaluation of the HTC-NN submodels

Both MLPs show promising results compared to measurements and especially compared to the numerical model SUEWS. With an overall (spatial) R^2 of 0.997 and an RMSE of 0.5 K, the T_a MLP shows excellent accuracy. Both SUEWS and the T_a

20

MLP have an RMSE of around 1.5 K in relation to sensor network data, while the MAE is only at around 1 K. RMSE also varies with time (Fig. 4c). Apart from the two periods late October and mid-December, with RMSE values higher than 2 K, RMSE fluctuates around 1 K for most of the remaining time. This suggests that the overall RMSE of 1.5 K is highly influenced by the two outlier periods in late October and mid-December with high RMSE values between 2 K and 3 K. For the RH MLP, similar observations are made, having higher RMSE in late October and mid-December. Since the MLPs and the SUEWS model have similar error patterns, the error must already be apparent in forcing data (see also Fig. B1Figure B1). The end of October was unusually warm, with clear sunny days but cold nights. As forcing data is measured on a rooftop at 55 m a.g.l., the near-surface night cooling is not represented in this data, leading to higher RMSE values. In December, on the other hand, an uncommon weather event occurred with a surface inversion between the surface and the rooftop at 55 m a.g.l., with colder air near the surface and warmer air above. Compared to the ERA5 land data, the forcing and model data show higher errors during these periods in October and December, indicating that errors are already present in the forcing data and are passed on to the model results. Both SUEWS and the MLP have a lower RMSE than the forcing data in relation to the sensor network data. However, the gain in accuracy is moderate for RH at 1 % and for T_a at 0.3 K. One reason could be that SUEWS is set up in default mode, and no specific model calibrations are performed. In addition, the modeled soil moisture has not been validated against measurement data. Another reason could be that SUEWS is run in offline mode. Running SUEWS in online mode, coupled to a mesoscale weather model would probably increase its accuracy further as boundary conditions and tiles interactions, including local wind systems, would be better mapped.

The RF approach to model a statistical wind field for the entire urbanized area of Freiburg shows good results for outdoor thermal comfort modeling. Since the RF models only a statistical wind field, not every wind gust can be accurately represented. Still, as, the model is forced with hourly aggregated data, it is sufficient to predict hourly averaged U.

The results of the evaluation of the T_a and RH MLPs and the RF illustrate the power and potential of deep and machine learning models in the context of urban climate when appropriate training data is available. It can also be seen, that the MLPs have similar model shortcomings as SUEWS, and an enhanced MLP performance requires an improved numerical model.

5.2   Evaluation HTC-NN

The tradeoff between computational cost and model accuracy of the HTC-NN compared to numerical models is positive for both. The HTC-NN has a higher model accuracy than SOLWEIG with its POI option. In addition, model accuracy is constant throughout the year, with lower accuracy during less common weather conditions. As mentioned, uncommon weather situations with strong surface inversions in the city are hard to predict for numerical and deep learning models. The overall RMSE of UTCI is around 3 K. The lower RMSE of the HTC-NN compared to SOLWEIG can be explained by the combination of four submodels that downscale T_a, RH, T_{mrt}, and U separately, while SOLWEIG downscales only T_{mrt} comprehensively. Diurnal error patterns of UTCI show higher errors during the day than during the night due to higher errors of T_{mrt} predictions. However, the highest errors occur before noon, while in the late afternoon, UTCI predictions are more accurate (RMSE of about 3K) when thermal stress is highest. During the heat wave event, the HTC-NN overestimates UTCI to different degrees

during the day. It is also apparent that daytime UTCI follows the patterns of T_{mrt} most of the time, emphasizing the importance of correct shadow pattern data. Since T_{mrt} predictions of SOLWEIG and HTC-NN are partly in line but UTCI values differ, the prediction error must be related not only to T_{mrt}, but to T_a and RH, while U is negligible during these days. These results indicate that the accuracy of the HTC-NN is affected to varying degrees by its submodels in different weather conditions and that an overall attribution of error to the submodels cannot be made but must be done individually for the different weather conditions.

To the authors' knowledge, not many comparisons have been made between UTCI models and measurements with at least one week of comparison. Nice et al. (2018) used a modified version of the TUF-3D model (Krayenhoff and Voogt, 2007) to model UTCI in a suburb of Melbourne for four weeks. Observed UTCI values were calculated from T_{mrt} (black globe), T_a, RH, and U measurements. A total of seven stations were installed and compared to model results. MAE between modeled and observed UTCI ranged from 1.80 K to 3.03 K and RMSE from 2.33 to 3.64 K, which aligns with HTC-NN. On the other hand, Meili et al. (2021) applied the ecohydrological model UT&C in Singapore with a model accuracy for UTCI ranging from 1.9 to 3.1 K (RMSE). Since Freiburg is hardly comparable to Melbourne or Singapore, these findings still help to better evaluate the HTC-NN model results. The HTC-NN has a similar accuracy but allows for modeling larger domains on high resolution.

While the HTC-NN has a very good trade-off between accuracy and computational cost, it has some limitations. First, the HTC-NN is not coupled to a mesoscale model and thus does not include local or mesoscale weather phenomena, such as mountain valley wind systems, cold air drainage and advection of heat (e.g., heat island and urban plume). Additionally, while large model domains are possible, the 500 x 500 m model tiles for T_a and RH are modeled individually, and no boundary or heat and moisture transport effects are considered, similar to the offline SUEWS version. These constraints are related to the model structure of the offline SUEWS version and could be partially resolved by running it coupled to a mesoscale model. Another constraint is related to the forcing data. The HTC-NN is forced with meteorological data from a weather station at 55 m a.g.l. Measured T_a at this height may already be exposed to other processes as near-ground level T_a, which may lead to initial error propagation. Additionally, the larger the model domain, the more difficult it is to force the model with data from one weather station. This could be solved by forcing the HTC-NN with reanalysis data or weather forecast products, which would reduce the model fit due to inconsistencies between measured and reanalysis and forecast data. As the submodels perform well in emulating the numerical models, better numerical models would also be needed to increase the accuracy of the model itself (Briegel et al. 2023). Nevertheless, the HTC-NN should only be applied to 'known' spatial and temporal data, as ANN are generally capable of interpolation but not extrapolation. This means that similar urban structures and/or meteorological forcing data are suitable as potential prediction data. However, any unknown spatial configurations or unknown extreme weather events, should be approached with caution and undergo validation against measurement or numerical model data.

The evaluation of the HTC-NN on sensor network data and the comparison with similar studies show that the HTC-NN is a valuable tool for modeling outdoor thermal comfort in complex urban areas.

5.3 High-resolution UTCI mapping

385 HTC-NN is used to predict UTCI for Freiburg for four years with high temporal (hourly) and spatial resolution (1 m). Almost 1.5 trillion predictions were made, which took around 8 days. To determine day and night heat hotspots, the hours above the specific 80th percentile (32 °C and 22 °C) are summed up and mapped. In addition, four specific areas representing LCZ 2, 5, and 8 are presented in more detail. The four areas differ in terms of their physical characteristics and also strongly in terms of their UTCI distributions. The daytime UTCI distributions of the new building district (d) and the industrial area (b) are flatter

390 and shifted to higher values compared to the distribution of the city center (a) and the old building district (c). The 80th percentiles are shifted up to 3 °C. The new building district and the industrial area also have more pixels with UTCI values ≥ 32 °C during the day. At night, however, it is the other way around, and the city center and the old building district have more pixels with UTCI > 22 °C, which is also present in the nighttime UTCI distributions. This effect can be largely attributed to the proportion of covered areas, either by tree canopy or dense building structure. The denser the buildings and the larger the

395 trees, the lower the sky view factor. Since T_{mrt} is the driving factor for outdoor thermal comfort during the day, a reduced SVF reduces the radiation load and, thus, the UTCI. At night, however, covered areas show reduced upward longwave radiation, leading to higher UTCI values. These results indicate that adaption measures may work inconsistently during the day and at night.

Besides the intra-urban variability of outdoor thermal comfort between different urban areas / LCZs, there is also an intra-

400 LCZ variability of outdoor thermal comfort. The old and new building districts have similar building characteristics but show very different UTCI distributions for day and night. Intra-LCZ variability has the same magnitude as inter-LCZ variability, as seen by their distributions and 80th percentiles. This is due to the remaining land cover characteristics besides building structures such as tree or grass cover fraction. Outdoor thermal comfort has high variability at the microscale due to U and T_{mrt}, highlighting the importance of high-resolution modeling. These results further indicate that the urban climate

405 characteristics of a small city with large green areas, such as Freiburg, are better described by the LCZ subclasses or directly by land cover class fractions.

6 Conclusions

This study presents a novel deep learning approach to outdoor thermal comfort modeling, the Human Thermal Comfort Neural Network (HTC-NN). The HTC-NN consists of four submodels that separately model T_a, RH, T_{mrt}, and U, followed by UTCI

410 calculation. The submodels are trained on numerical model results, essentially emulating the numerical models through machine learning methods. Each submodel and the final HTC-NN are validated separately with data from a dense sensor network.

The research objective to develop and evaluate machine learning models emulating numerical urban climate models could be achieved (i). In addition, the evaluation of the HTC-NN shows promising results (ii). Furthermore, we could show that the

415 HTC-NN has a positive trade-off between accuracy and computational cost. The accuracy of the HTC-NN is comparable to

numerical urban climate models (RMSE of 3 K to 3.3 K from SOLWEIG), while it is computationally superior. This computational superiority allows high-resolution modeling of outdoor thermal comfort for large domains and long periods. The HTC-NN is used to model UTCI for four years (2019–2022) for Freiburg with a temporal and spatial resolution of one hour and 1 x 1 m to determine intra-urban variability of outdoor thermal comfort (research objective iii). The advantage of the HTC-NN is that it is able to model outdoor thermal comfort with high spatial and temporal resolution, which allows the investigation of spatial and temporal patterns of outdoor thermal comfort. Therefore, high resolution UTCI predictions can be aggregated either temporally or spatially, or both. The HTC-NN further allows studying thermal comfort patterns only for specific weather events, UTCI classes, or selected areas.

We demonstrate that HTC-NN is fast and versatile enough to continuously model long periods and entire cities using a building-resolved approach. For urban climate services, but also urban climate assessments and environmental consulting applications, there is no longer a need to base assessments on numerical simulations of a few selected "case studies", but instead we can build entire climatologies of thermal comfort, including the corresponding exceedance frequencies, explicitly on end-user machines.

Nevertheless, the HTC-NN has limitations, and its applicability to other cities needs further investigation. It does not consider the city surroundings and its mountains with elevation differences up to 1000 m, as it is not coupled to a mesoscale weather model. This means it does not consider local wind systems, shading by hills, or interactions between different local scale grid cells (500 x 500 m). In addition, heat advection is not taken into account. Regarding the transferability to other cities, as long as the city has similar building and vegetation characteristics, it could be easily transferred since the training data covers a wide range of urban structures. However, mapping unknown urban forms, such as skyscrapers or denser building structures, is critical, as they are not present in the training data. Nevertheless, it could be applied to unknown cities after validation of the trained submodels by numerical models or on measured data.

Although only one potential application of the HTC-NN has been considered in the present paper, several applications are enabled by the low computational cost of the HTC-NN. These applications include "simple" urban-specific thermal comfort predictions and warnings for the next days using weather forecast models. The HTC-NN, however, could also be applied to the downscaling of potential future climates to building scale (e.g., using EURO-CORDEX data), or the assessment of the effectiveness of adaptation measures by changing the input data on urban form. Finally, HTC-NN can be used to investigate the driving forces of thermal comfort at different scales and hence fundamentally develop guidelines in support of urban planning and policymaking.

Code and data availability

The code of the MLPs, the U-Net, and the UTCI calculation are available at https://doi.org/10.5281/zenodo.7974472. However, the digital elevation model data are released by the city of Freiburg only on a restricted basis. Nevertheless, they can be

requested for scientific purposes from the Surveying Office of the city of Freiburg. All spatial and meteorological data and model results can be found at https://doi.org/10.5281/zenodo.7974307.

Author contributions

450 FB designed the HTC-NN and prepared the original draft in cooperation with AC. JW designed the RF wind model. FB analysed and visualised the results. AC and DS provided supervision and review of the original draft.

Competing interests

The authors declare that they have no conflict of interest.

Appendix A

455 **Table A1: Universal thermal climate index (UTCI) classification of thermal stress.**

| UTCI (°C)	Stress category
> +46	Extreme heat stress
+38 – +46	Very strong heat stress
+32 – +36	Strong heat stress
+26 – +32	Moderate heat stress
+9 – +26	No thermal stress
0 – +9	Slight cold stress
-13 – 0	Moderate cold stress
-27 – -13	Strong cold stress
-40 – -27	Very strong cold stress
< -40	Extreme cold stress
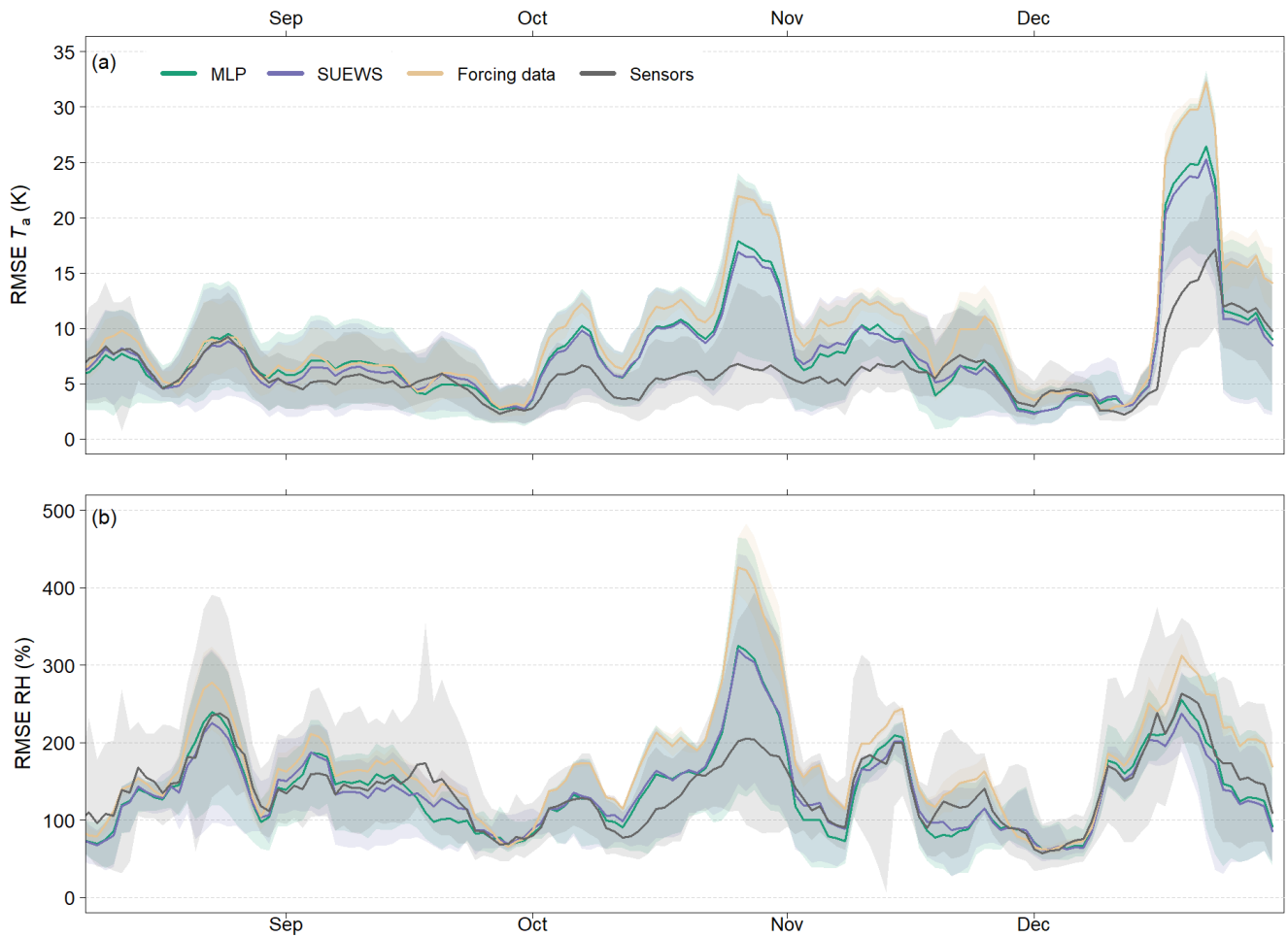
Figure B1: Moving average of RMSE of T_a (a) and RH (b) from August to December 2022. As reference data ERA5-Land data is used (Muñoz Sabater, 2019). The window size of the moving average is seven days. Time series starts with the installation of the first Tier I stations in August 2022. Shaded areas represent 95 % confidence interval.

Table B1: Validation measures of the machine learning models against the numerical models. In contrast to 4, where only pointwise comparisons are made, this table shows mean, MAE, MBE, RMSE, and R^2 for the entire test areas (spatial comparison) for 2022.

Variable	Model combination	Mean	MAE	MBE	RMSE	R^2
T_a	SUEWS / MLP	13.36 °C / 13.38 °C	0.31 K	-0.02 K	0.50 K	0.997
RH	SUEWS / MLP	70.43 % / 70.69 %	2.14 %	-0.26 %	3.28 %	0.97
U	LES / RF	1.12 m s^{-1} / 0.58 m s^{-1}	0.55 m s^{-1}	-0.54 m s^{-1}	0.76 m s^{-1}	0.30
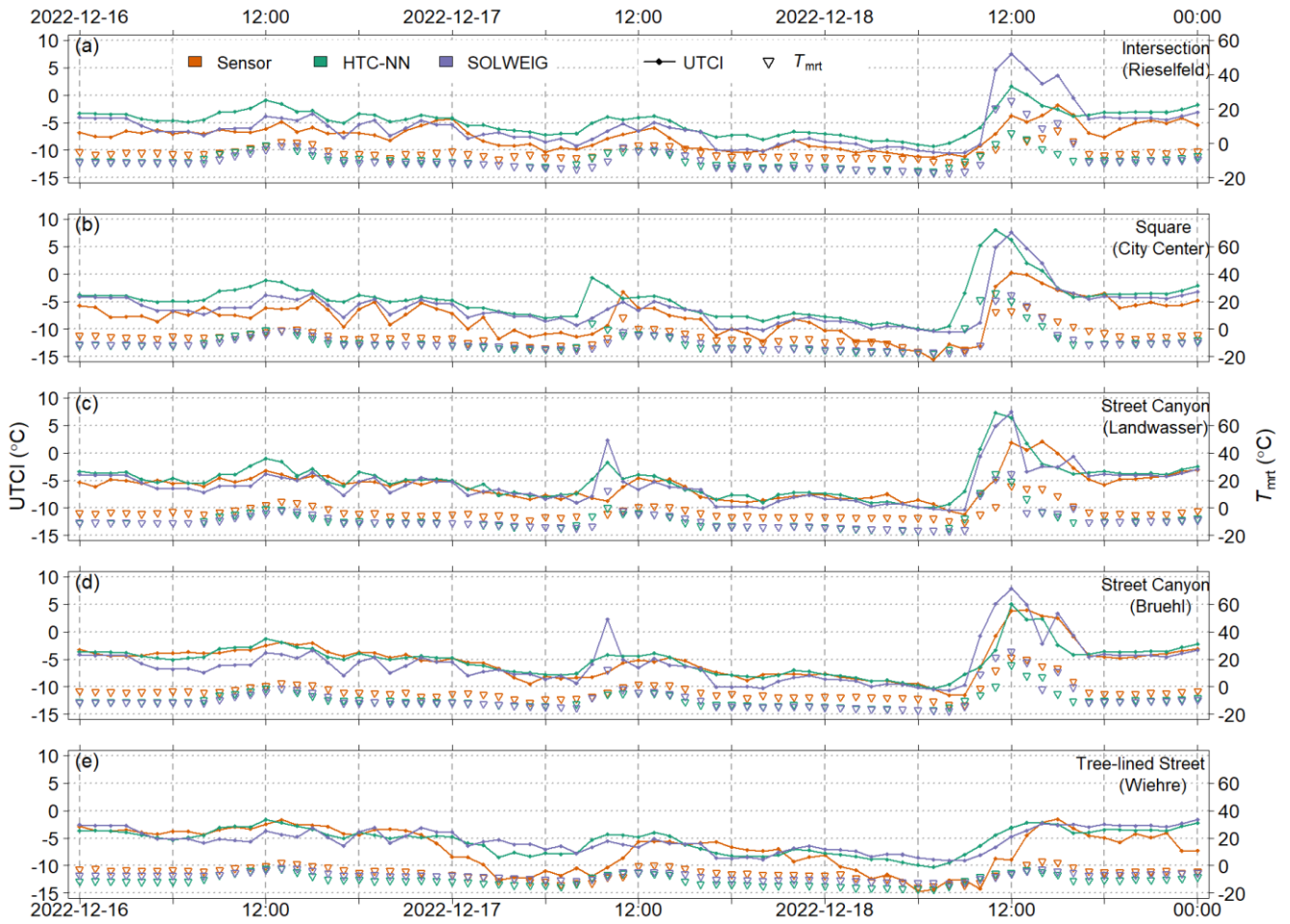
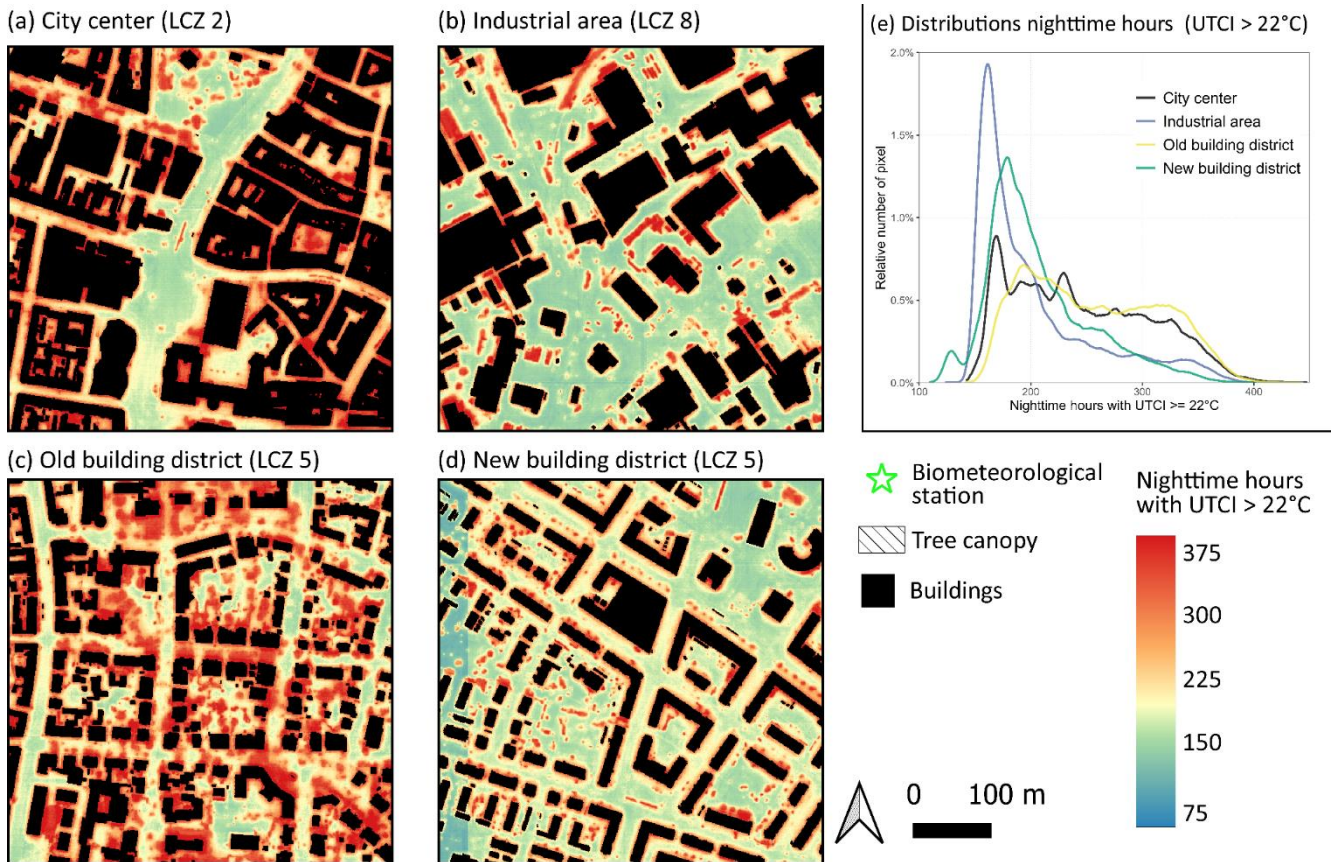**Figure B2: Modeled and measured UTCI at five Tier I stations (a)–(e) during a cold wave in December 2022.**

Figure B3: Four 500 x 500m subsets of different urban neighborhoods of the city of Freiburg. Average nighttime hours with a UTCI ≥ 22 °C are shown. (a) shows the city center of Freiburg, (b) an industrial area in the north of Freiburg, (c) a district with old buildings and dense / mature tree stock, and (d) a building district built after the year 1995. (e) shows the distributions of (a)– (e). Note, coloring is in accordance with predicted UTCI quantiles. Spatial resolution is 1 x 1m and time period is 2019–2022.

465

475 **References**

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A Next-Generation Hyperparameter Optimization Framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2623–2631, https://doi.org/10.1145/3292500.3330701, 2019.

Albertson, J. D. and Parlange, M. B.: Natural integration of scalar fluxes from complex terrain, Adv Water Resour, 23, 239–252, https://doi.org/10.1016/S0309-1708(99)00011-1, 1999a.

Albertson, J. D. and Parlange, M. B.: Surface length scales and shear stress: Implications for land-atmosphere interaction over complex terrain, Water Resour Res, 35, 2121–2132, https://doi.org/10.1029/1999WR900094, 1999b.

Ao, X., Grimmond, C. S. B., Ward, H. C., Gabey, A. M., Tan, J., Yang, X.-Q., Liu, D., Zhi, X., Liu, H., and Zhang, N.: Evaluation of the Surface Urban Energy and Water Balance Scheme (SUEWS) at a Dense Urban Site in Shanghai: Sensitivity to Anthropogenic Heat and Irrigation, J Hydrometeorol, 19, 1983–2005, https://doi.org/10.1175/JHM-D-18-0057.1, 2018.

Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B.: Algorithms for Hyper-Parameter Optimization, in: Proceedings of the 24th International Conference on Neural Information Processing Systems, 2546–2554, 2011.

Bergstra, J., Yamins, D., and Cox, D.: Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures, in: Proceedings of the 30th International Conference on Machine Learning, 115–123, 2013.

Best, M. J. and Grimmond, C. S. B.: Key Conclusions of the First International Urban Land Surface Model Comparison Project, Bull Am Meteorol Soc, 96, 805–819, https://doi.org/10.1175/BAMS-D-14-00122.1, 2015.

Błażejczyk, K., Jendritzky, G., Bröde, P., Fiala, D., Havenith, G., Epstein, Y., Psikuta, A., and Kampmann, B.: An introduction to the Universal Thermal Climate Index (UTCI), 2013.

Briegel, F., Lee, S. C., Black, T. A., Jassal, R. S., and Christen, A.: Factors controlling long-term carbon dioxide exchange between a Douglas-fir stand and the atmosphere identified using an artificial neural network approach, Ecol Modell, 435, 109266, https://doi.org/10.1016/j.ecolmodel.2020.109266, 2020.

Briegel, F., Makansi, O., Brox, T., Matzarakis, A., and Christen, A.: Modelling long-term thermal comfort conditions in urban environments using a deep convolutional encoder-decoder as a computational shortcut, Urban Clim, 47, 101359, https://doi.org/10.1016/j.uclim.2022.101359, 2023.

Chen, F., Kusaka, H., Bornstein, R., Ching, J., Grimmond, C. S. B., Grossman-Clarke, S., Loridan, T., Manning, K. W., Martilli, A., Miao, S., Sailor, D., Salamanca, F. P., Taha, H., Tewari, M., Wang, X., Wyszogrodzki, A. A., and Zhang, C.: The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems, International Journal of Climatology, 31, 273–288, https://doi.org/10.1002/joc.2158, 2011.

Chen, S., Yang, Y., Deng, F., Zhang, Y., Liu, D., Liu, C., and Gao, Z.: A high-resolution monitoring approach of canopy urban heat island using a random forest model and multi-platform observations, Atmos Meas Tech, 15, 735–756, https://doi.org/10.5194/amt-15-735-2022, 2022.

Coccolo, S., Kämpf, J., Scartezzini, J.-L., and Pearlmutter, D.: Outdoor human comfort and thermal stress: A comprehensive review on models and standards, Urban Clim, 18, 33–57, https://doi.org/10.1016/j.uclim.2016.08.004, 2016.

Cohen, P., Potchter, O., and Matzarakis, A.: Daily and seasonal climatic conditions of green urban open spaces in the Mediterranean climate and their impact on human comfort, Build Environ, 51, 285–295, https://doi.org/https://doi.org/10.1016/j.buildenv.2011.11.020, 2012.

CRED: Cred Crunch 62 - 2020 Annual Report, Centre for Research on the Epidemiology of Disasters, 2021.

Demuzere, M., Kittner, J., Martilli, A., Mills, G., Moede, C., Stewart, I. D., van Vliet, J., and Bechtel, B.: A global map of local climate zones to support earth system modelling and urban-scale environmental science, Earth Syst Sci Data, 14, 3835–3873, https://doi.org/10.5194/essd-14-3835-2022, 2022.

Dupont, S., Mestayer, P. G., Guilloteau, E., Berthier, E., and Andrieu, H.: Parameterization of the Urban Water Budget with the Submesoscale Soil Model, J Appl Meteorol Climatol, 45, 624–648, https://doi.org/10.1175/JAM2363.1, 2006.

Epstein, Y. and Moran, D. S.: Thermal Comfort and the Heat Stress Indices, Ind Health, 44, 388–398, https://doi.org/10.2486/indhealth.44.388, 2006.

Feigel, G., Plein, M., Zeeman, M., Briegel, F., and Christen, A.: A compact and customisable street-level sensor system for real-time weather monitoring and outreach in Freiburg, Germany, EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-15609, https://doi.org/10.5194/egusphere-egu23-15609, 2023.

Fenner, D., Meier, F., Bechtel, B., Otto, M., and Scherer, D.: Intra and inter 'local climate zone' variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany, Meteorologische Zeitschrift, 26, 525–547, https://doi.org/10.1127/metz/2017/0861, 2017.

City of Freiburg im Breisgau - Bevölkerung: https://www.freiburg.de/pb/site/Freiburg/node/207904?QUERYSTRING=Stadtbezirk%20Wohnbevoelkerung, last access: 22 June 2022.

Giometto, M. G., Christen, A., Meneveau, C., Fang, J., Krafczyk, M., and Parlange, M. B.: Spatial Characteristics of Roughness Sublayer Mean Flow and Turbulence Over a Realistic Urban Surface, Boundary Layer Meteorol, 160, 425–452, https://doi.org/10.1007/s10546-016-0157-6, 2016.

Giometto, M. G., Christen, A., Egli, P. E., Schmid, M. F., Tooke, R. T., Coops, N. C., and Parlange, M. B.: Effects of trees on mean wind, turbulence and momentum exchange within and above a real urban environment, Adv Water Resour, 106, 154–168, https://doi.org/10.1016/j.advwatres.2017.06.018, 2017.

Grimmond, C. S. B., Blackett, M., Best, M. J., Baik, J.-J., Belcher, S. E., Beringer, J., Bohnenstengel, S. I., Calmet, I., Chen, F., Coutts, A., Dandou, A., Fortuniak, K., Gouvea, M. L., Hamdi, R., Hendry, M., Kanda, M., Kawai, T.,

540   Kawamoto, Y., Kondo, H., Krayenhoff, E. S., Lee, S.-H., Loridan, T., Martilli, A., Masson, V., Miao, S., Oleson, K., Ooka, R., Pigeon, G., Porson, A., Ryu, Y.-H., Salamanca, F., Steeneveld, G. J., Tombrou, M., Voogt, J. A., Young, D. T., and Zhang, N.: Initial results from Phase 2 of the international urban energy balance model comparison, International Journal of Climatology, 31, 244–272, https://doi.org/10.1002/joc.2227, 2011.

Gubler, M., Christen, A., Remund, J., and Brönnimann, S.: Evaluation and application of a low-cost measurement
545   network to study intra-urban temperature differences during summer 2018 in Bern, Switzerland, Urban Clim, 37, 100817, https://doi.org/10.1016/j.uclim.2021.100817, 2021.

Hamdi, R., Degrauwe, D., and Termonia, P.: Coupling the Town Energy Balance (TEB) Scheme to an Operational Limited-Area NWP Model: Evaluation for a Highly Urbanized Area in Belgium, Weather Forecast, 27, 323–344, https://doi.org/10.1175/WAF-D-11-00064.1, 2012.

550   Hamdi, R., Kusaka, H., Doan, Q.-V., Cai, P., He, H., Luo, G., Kuang, W., Caluwaerts, S., Duchêne, F., van Schaeybroek, B., and Termonia, P.: The State-of-the-Art of Urban Climate Change Modeling and Observations, Earth Systems and Environment, 4, 631–646, https://doi.org/10.1007/s41748-020-00193-3, 2020.

Ho, H. C., Knudby, A., Sirovyak, P., Xu, Y., Hodul, M., and Henderson, S. B.: Mapping maximum urban air temperature on hot summer days, Remote Sens Environ, 154, 38–45, https://doi.org/10.1016/j.rse.2014.08.012, 2014.

555   Holst, J. and Mayer, H.: Impacts of street design parameters on human-biometeorological variables, Meteorologische Zeitschrift, 20, 241–552, https://doi.org/10.11270941-294820110254, 2011.

IPCC: Index, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy,
560   E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, In press, 2021.

Järvi, L., Grimmond, C. S. B., and Christen, A.: The Surface Urban Energy and Water Balance Scheme (SUEWS): Evaluation in Los Angeles and Vancouver, J Hydrol (Amst), 411, 219–237, https://doi.org/10.1016/j.jhydrol.2011.10.001, 2011.

565   Kanda, M., Kanega, M., Kawai, T., Moriwaki, R., and Sugawara, H.: Roughness Lengths for Momentum and Heat Derived from Outdoor Urban Scale Models, J Appl Meteorol Climatol, 46, 1067–1079, https://doi.org/10.1175/JAM2500.1, 2007.

Kántor, N. and Unger, J.: The most problematic variable in the course of human-biometeorological comfort assessment — the mean radiant temperature, Open Geosciences, 3, 90–100, https://doi.org/doi:10.2478/s13533-011-0010-x, 2011.

570   Krayenhoff, E. S. and Voogt, J. A.: A microscale three-dimensional urban energy balance model for studying surface temperatures, Boundary Layer Meteorol, 123, 433–461, https://doi.org/10.1007/s10546-006-9153-6, 2007.

Lee, H., Holst, J., and Mayer, H.: Modification of Human-Biometeorologically Significant Radiant Flux Densities by Shading as Local Method to Mitigate Heat Stress in Summer within Urban Street Canyons, Advances in Meteorology, 2013, 312572, https://doi.org/10.1155/2013/312572, 2013.

575 Lindberg, F. and Grimmond, C. S. B.: SOLWEIG_v2019a, Department of Earth Sciences, University of Gothenburg, Sweden, University of Reading, UK., 2019.

Martilli, A., Clappier, A., and Rotach, M. W.: An Urban Surface Exchange Parameterisation for Mesoscale Models, Boundary Layer Meteorol, 104, 261–304, https://doi.org/10.1023/A:1016099921195, 2002.

Masson, V.: A Physically-Based Scheme For The Urban Energy Budget In Atmospheric Models, Boundary Layer

580 Meteorol, 94, 357–397, https://doi.org/10.1023/A:1002463829265, 2000.

Masson, V., Lemonsu, A., Hidalgo, J., and Voogt, J.: Urban Climates and Climate Change, Annu Rev Environ Resour, 45, 411–444, https://doi.org/10.1146/annurev-environ-012320-083623, 2020.

Matzarakis, A., Martinelli, L., and Ketterer, C.: Relevance of Thermal Indices for the Assessment of the Urban Heat Island, in: Counteracting Urban Heat Island Effects in a Global Climate Change Scenario, edited by: Musco, F., Springer

585 International Publishing, Cham, 93–107, https://doi.org/10.1007/978-3-319-10425-6_4, 2016.

Meili, N., Acero, J. A., Peleg, N., Manoli, G., Burlando, P., and Fatichi, S.: Vegetation cover and plant-trait effects on outdoor thermal comfort in a tropical city, Build Environ, 195, 107733, https://doi.org/10.1016/j.buildenv.2021.107733, 2021.

Meyer, D., Grimmond, S., Dueben, P., Hogan, R., and van Reeuwijk, M.: Machine Learning Emulation of Urban Land

590 Surface Processes, J Adv Model Earth Syst, 14, e2021MS002744, https://doi.org/10.1029/2021MS002744, 2022.

Mirzaei, P. A.: Recent challenges in modeling of urban heat island, Sustain Cities Soc, 19, 200–206, https://doi.org/10.1016/j.scs.2015.04.001, 2015.

Nice, K. A., Coutts, A. M., and Tapper, N. J.: Development of the VTUF-3D v1.0 urban micro-climate model to support assessment of urban vegetation influences on human thermal comfort, Urban Clim, 24, 1052–1076,

595 https://doi.org/10.1016/j.uclim.2017.12.008, 2018.

Plein, M., Feigel, G., Zeeman, M., Briegel, F., Dormann, C., and Christen, A.: A sensor network for real-time monitoring and modelling of street-level heat exposure in Freiburg, Germany, EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-13816, https://doi.org/10.5194/egusphere-egu23-13816, 2023.

Potchter, O., Cohen, P., Lin, T.-P., and Matzarakis, A.: Outdoor human thermal perception in various climates: A

600 comprehensive review of approaches, methods and quantification, Science of The Total Environment, 631–632, 390–406, https://doi.org/10.1016/j.scitotenv.2018.02.276, 2018.

Quanz, J. A., Ulrich, S., Fenner, D., Holtmann, A., and Eimermacher, J.: Micro-Scale Variability of Air Temperature within a Local Climate Zone in Berlin, Germany, during Summer, Climate, 6, https://doi.org/10.3390/cli6010005, 2018.

32

Rafael, S., Martins, H., Matos, M. J., Cerqueira, M., Pio, C., Lopes, M., and Borrego, C.: Application of SUEWS model forced with WRF: Energy fluxes validation in urban and suburban Portuguese areas, Urban Clim, 33, 100662, https://doi.org/10.1016/j.uclim.2020.100662, 2020.

Rousi, E., Kornhuber, K., Beobide-Arsuaga, G., Luo, F., and Coumou, D.: Accelerated western European heatwave trends linked to more-persistent double jets over Eurasia, Nat Commun, 13, 3851, https://doi.org/10.1038/s41467-022-31432-y, 2022.

Shreevastava, A., Prasanth, S., Ramamurthy, P., and Rao, P. S. C.: Scale-dependent response of the urban heat island to the European heatwave of 2018, Environmental Research Letters, 16, 104021, https://doi.org/10.1088/1748-9326/ac25bb, 2021.

Staiger, H., Laschewski, G., and Matzarakis, A.: Selection of Appropriate Thermal Indices for Applications in Human Biometeorological Studies, https://doi.org/10.3390/atmos10010018, 2019.

Stewart, I. D. and Oke, T. R.: Local Climate Zones for Urban Temperature Studies, Bull Am Meteorol Soc, 93, 1879–1900, https://doi.org/10.1175/BAMS-D-11-00019.1, 2012.

Straub, A., Berger, K., Breitner, S., Cyrys, J., Geruschkat, U., Jacobeit, J., Kühlbach, B., Kusch, T., Philipp, A., Schneider, A., Umminger, R., Wolf, K., and Beck, C.: Statistical modelling of spatial patterns of the urban heat island intensity in the urban environment of Augsburg, Germany, Urban Clim, 29, 100491, https://doi.org/10.1016/j.uclim.2019.100491, 2019.

Sun, T. and Grimmond, S.: A Python-enhanced urban land surface model SuPy (SUEWS in Python, v2019.2): development, deployment and demonstration, Geosci Model Dev, 12, 2781–2795, https://doi.org/10.5194/gmd-12-2781-2019, 2019.

Sun, T., ljarvi, Omidvar, H., LewisB7, natalieth, biglimp, Li, Z., Grimmond, S., and pjaysuews: UMEP-dev/SUEWS: 2020a Release, https://doi.org/10.5281/zenodo.5723970, November 2021.

Unger, J., Skarbit, N., Kovács, A., and Gál, T.: Comparison of regional and urban outdoor thermal stress conditions in heatwave and normal summer periods: A case study, Urban Clim, 32, 100619, https://doi.org/10.1016/j.uclim.2020.100619, 2020.

Ward, H. C., Evans, J. G., and Grimmond, C. S. B.: Multi-season eddy covariance observations of energy, water and carbon fluxes over a suburban area in Swindon, UK, Atmos Chem Phys, 13, 4645–4666, https://doi.org/10.5194/acp-13-4645-2013, 2013.

Ward, H. C., Kotthaus, S., Järvi, L., and Grimmond, C. S. B.: Surface Urban Energy and Water Balance Scheme (SUEWS): Development and evaluation at two UK sites, Urban Clim, 18, 1–32, https://doi.org/10.1016/j.uclim.2016.05.001, 2016.