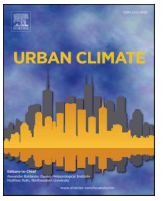
Urban Climate 57 (2024) 102114

Contents lists available at ScienceDirect

Urban Climate

journal homepage: www.elsevier.com/locate/uclim



High-resolution urban climate simulations for heat and health applications in Philadelphia

Benjamin Le Roy [a,*], Keith W. Dixon [b], Dennis Adams-Smith [c]

[a] Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey, USA
[b] National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA
[c] University Corporation for Atmospheric Research Cooperative Program for the Advancement of Earth System Science, Boulder, CO, USA

ARTICLE INFO

Keywords:
Urban Climate Modeling
SURFEX
TEB
Urban Heat Island
Evaluation
Heat Stress

ABSTRACT

Stakeholders need high-resolution urban climate information for city planning and adaptation to climate risks. Climate models have too coarse a spatial resolution to properly represent cities at the relevant scale, and downscaled products often fail to account for urban effects. We propose here a methodological framework for producing high-resolution urban databases that are used to drive the SURFEX-TEB land surface and urban canopy models. A historical simulation is carried out over the period 1991–2020, based on a reanalysis of the city of Philadelphia (Pennsylvania, USA). The simulation is compared with observations outside and inside the city, as well as with a field campaign. The results show good agreement between the model and observations, with average summer biases of only −1 °C and + 0.8 °C for daily minimum and maximum temperatures outside the city, and almost none inside. The simulation is used to calculate the maximum daily heat index (HIX) and to study emergency heat alerts. The HIX is slightly overestimated and, consequently, the model simulates too many heat events if not bias corrected. Overall, HIX conditions at Philadelphia International Airport are found to be suitable proxies for city-wide summer conditions, and therefore are appropriate to use for emergency heat declarations.

1. Introduction

Most of the United States' population lives in cities and is subjected to health impacts specific to the urban environment (heat stress Heaviside et al., 2017, Chu et al., 2023, Hayden et al., 2023; air pollution Brugha and Grigg, 2014). Because of their geometry, the materials they are made of, and the activities they host, cities modify their local environment (Oke et al., 2017) with the Urban Heat Island (UHI, Arnfield, 2003; Rizwan et al., 2008) being the most striking effect. Those impacts have been intensifying because of regional climate change and urbanization (Argüeso et al., 2015; Oleson et al., 2015; D'Amato et al., 2010). Stakeholders and local decision makers need relevant high-resolution urban climate information to properly assess areas that are more vulnerable to those hazards and therefore can benefit from scientifically-informed climate adaptation measures (Masson et al., 2014, 2020).

The global and regional dynamical (physics-based) climate models that are used to produce regional climate simulations (Doblas-Reyes et al., 2021; Ranasinghe et al., 2021) on climatological time scales (i.e. more than a few decades) typically are not well suited for studying the urban scale because their horizontal resolution is too coarse and they do not simulate important features of the urban environment, such as its geometry or certain physical processes like anthropogenic heat (Langendijk et al., 2019; Le Roy et al., 2021).

* Corresponding author at: Climate Service Center Germany (GERICS), Helmholtz-Zentrum Hereon, Hamburg, Germany.
E-mail address: benjamin.le-roy@hereon.de (B. Le Roy).

https://doi.org/10.1016/j.uclim.2024.102114
Received 3 May 2024; Received in revised form 18 July 2024; Accepted 21 August 2024
Available online 29 August 2024

Because of their high computational costs, higher spatial resolution dynamical model projections are often limited to a few scenarios and limited simulation lengths. Statistical downscaling methodologies exist to bridge the resolution gap, however their application over urban areas is limited by the lack of reliable long observation records representative of urban microclimates.

Previous studies have simulated the local climate of Philadelphia, Pennsylvania, USA using some of the approaches outlined above, but often for a few days only. Otte et al. (2004) evaluated the added value of an Urban Canopy Model (UCM) in the Penn State–NCAR fifth-generation Mesoscale Model (MM5; Grell et al., 1994), with Philadelphia as a case study. Using the same model, Fan and Sailor (2005) studied the impacts of anthropogenic heat on urban air temperature over 2 days in winter and summer. Baniassadi et al. (2019) used the EnergyPlus model to quantify the effects of roof materials for a Typical Meteorological Year. Hashemi et al. (2023) simulated the thermal comfort of two building blocks in the city using the ENVI-Met (Bruse and Fleer, 1998; Bruse, 2004) model coupled to Bio-Met for 6 different days representing varying climatic conditions (seasons and extreme cold and heat). Trail et al. (2013) dynamically downscaled NASA earth system GISS ModelE2 (Schmidt et al., 2006, 2014) using the mesoscale atmospheric model WRF (Skamarock and Klemp, 2008) to study the impacts of climate change on urban air quality without the explicit representation of cities. Using the same model, but this time coupled to an UCM (the Princeton Urban Canopy Model, Wang et al., 2013), Ramamurthy and Bou-Zeid (2017) simulated the urban climate of New York and Philadelphia for 28 days surrounding a past heat wave.

A last avenue that has been applied successfully in the past in Europe (Lemonsu et al., 2012, 2013; Hamdi et al., 2015; Nogueira et al., 2020a, 2020b; Duchêne et al., 2022; Le Roy et al., 2024) and in the USA by Chakraborty et al. (2023) is the use of the Land Surface Model component (LSM) of these climate models coupled to Urban Canopy Models (UCM) in standalone mode (without simulating feedbacks with the atmosphere). These offline simulations make it possible to reproduce the urban climate at high spatial (under the kilometer scale) and temporal (sub-daily) resolutions for long time periods (more than a century), while producing relevant urban climate indicators, such as heat stress or energy consumption of buildings. This approach was recently used by Chakraborty et al. (2023) to study urban heat stress disparities in the United States using the High-Resolution Land Data Assimilation System (Chen et al., 2007) to drive the Weather Research and Forecasting model (Skamarock and Klemp, 2008) coupled to the Kusaka et al. (2001) single layer urban canopy model.

Here we use the LSM SURFEX (Masson et al., 2013) integrating the UCM TEB (Masson, 2000, Redon et al., 2020) driven by the global atmospheric ECMWF reanalysis ERA5 (Hersbach et al., 2020) to study the urban climate of Philadelphia (Fig. 1). The modeling framework is presented in detail in section 2 and the multi-decadal simulation is evaluated in section 3. Model results are used to study the heat conditions of the city with an emphasis on the Heat Index metric (Steadman, 1979; Anderson et al., 2013), which the city's Health Department refers to when issuing Heat Health Emergency declarations. Concluding remarks are presented in section 4.

2. Modeling framework

2.1. A land surface model for urban areas

To simulate the historical urban climate of Philadelphia, we use the openly available version of the land surface modeling platform

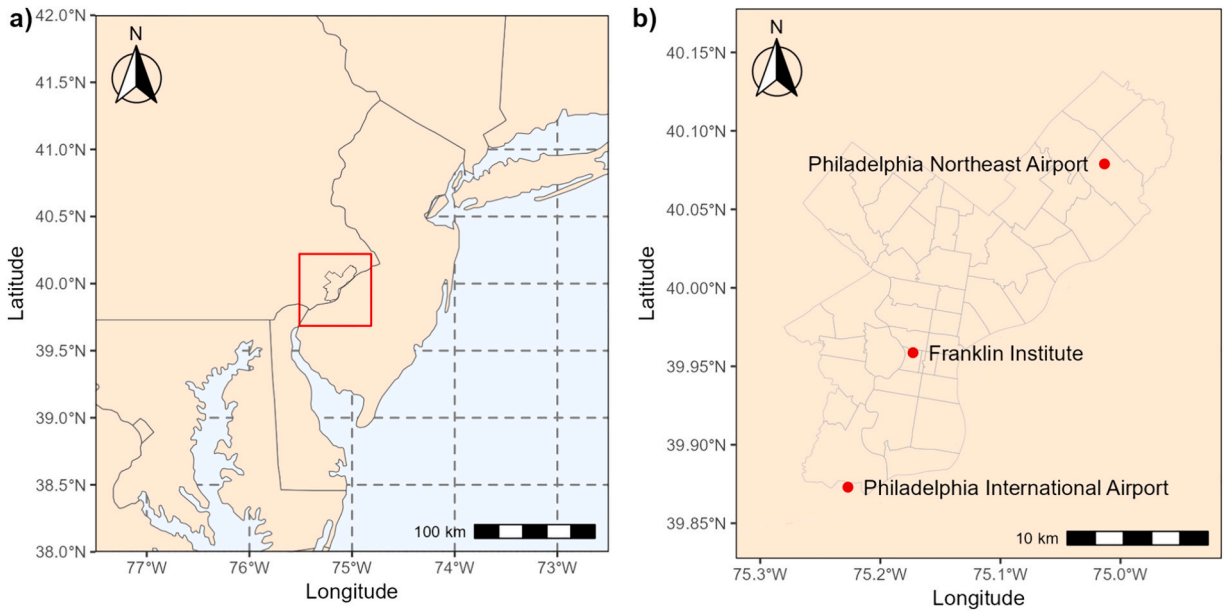


Fig. 1. a) Map of the northeast coast of the USA, where the study area is located. The red rectangle represents the SURFEX simulation domain, and the city limits of Philadelphia are also shown. b) Location of the three stations used for the model evaluation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

SURFEX (Masson et al., 2013), version 8.1. SURFEX represents different types of surfaces and uses a dedicated model to compute exchanges of heat, mass and momentum between each surface and the atmosphere. Natural vegetated areas are modeled by the Soil–Vegetation–Atmosphere-Transfer ISBA (Noilhan and Planton, 1989), the inland water bodies by the FLake model (Mironov, 2008; Salgado and Le Moigne, 2010), the seas and oceans by a one-dimensional oceanic mixing layer model, and the urban and artificial areas by the urban canopy model (UCM) TEB (Town Energy Balance, Masson, 2000). SURFEX is based on a fractional tiling approach: each model grid point has an associated proportion of each type of surface (i.e. natural vegetation and town) that is used to average the fluxes computed by each model component separately based on the same atmospheric conditions (either coming from a coupled

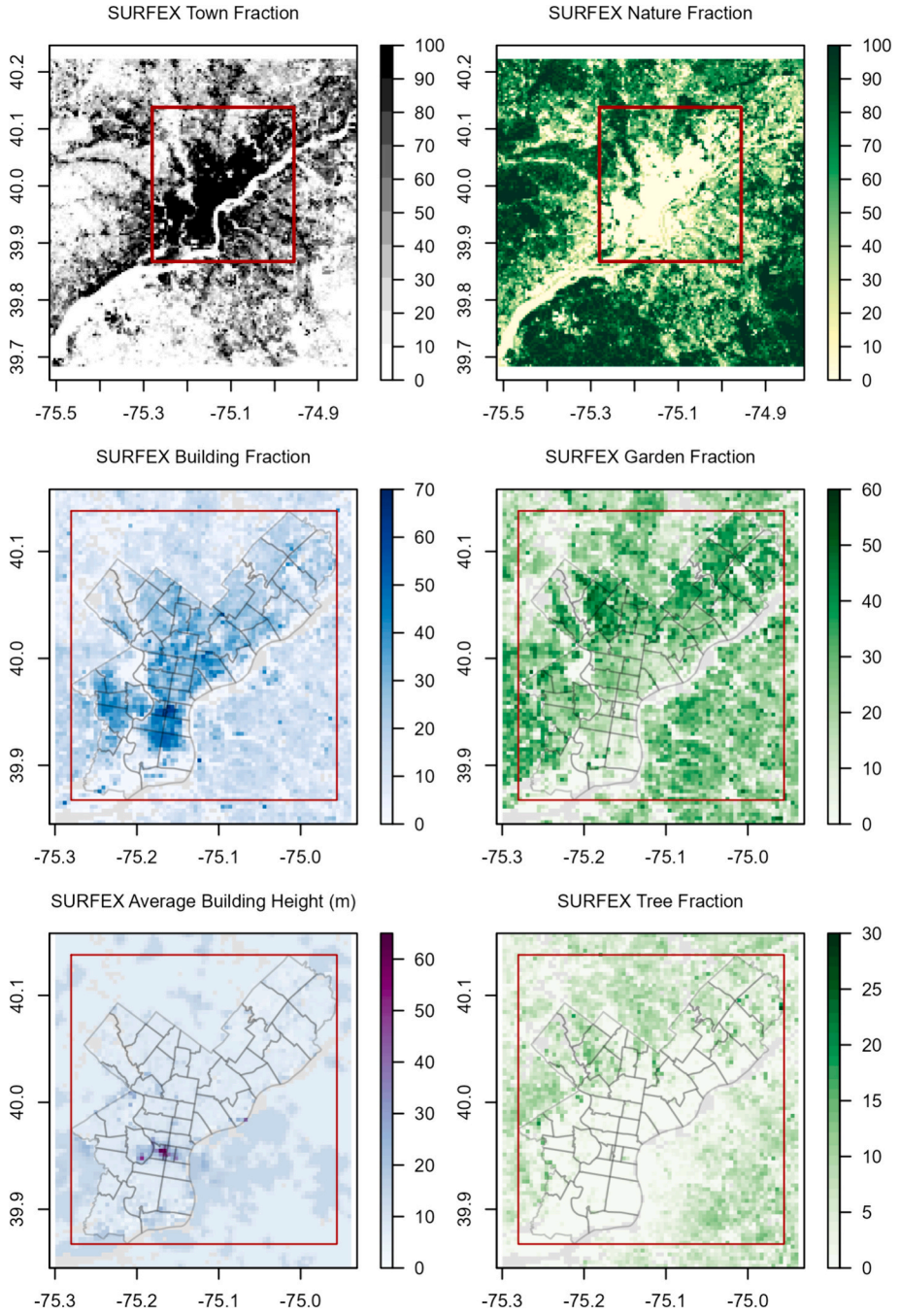


Fig. 2. Overview of the different land fractions defined on the SURFEX grid for the Philadelphia region. Note the upper two panels depict a larger area than the lower four. The red boxes represent the extent of the city limits for which another dataset has been used to improve the representation of fraction and building height. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

atmospheric model or prescribed by the user) described in Section 2.3. The SURFEX configuration used in this study has a 150 by 150 horizontal grid domain, with a grid cell resolution of 400 m (Fig. 2). Note that SURFEX does not represent horizontal exchanges between points, in effect assuming that the influence of horizontal advection is represented in the atmospheric forcing.

For our case study, what is of most interest is the representation of cities by TEB. TEB is a state-of-the-art UCM based on the urban canyon approach (Nunez and Oke, 1977). For each grid cell, the city fraction is simplified as a canyon composed of a roof, a wall and a road of infinite length and every possible orientation. TEB computes the energy balances of each surface, their exchanges with the atmosphere inside the canyon and then above it. Between its initial formulation and the version used here, TEB has been improved by numerous developments, such as the modeling of different atmospheric levels inside the urban canyon through a drag force approach (Hamdi and Masson, 2008), the representation of urban vegetation (Lemonsu et al., 2012) and rooftop vegetation (De Munck et al., 2013), and the integration of a Building Energy Model (BEM, Bueno et al., 2012). In this version, urban vegetation is modeled using the same module as for natural vegetation: ISBA, but called inside the canyon. This has the effect of changing the canyon geometry (e.g. extending its width with the presence of gardens), the fluxes received by the vegetation (shading of buildings, effects of walls) and consequently the temperature inside the city.

2.2. Surface boundary conditions

The SURFEX modeling platform requires several types of information to be specified for each grid cell, in order to characterize the surface to be modeled. The specified information includes, but is not limited to, the fraction of city, nature, inland water, sea and ocean within a grid cell, and physical parameters such as albedo, emissivity and heat capacity. By default, SURFEX refers to the global land cover database ECOCLIMAP (Champeaux et al., 2005) which maps 215 land cover and land use classes at a resolution of 1 km and provides an average value for all the necessary surface parameters. Though the use of ECOCLIMAP is appropriate for vegetated areas, outside Europe the database represents only one urban class with values closer to average suburban areas than to dense city centers. Therefore, to better represent the city of Philadelphia and to better simulate its climate, we provide SURFEX with an updated dataset at higher resolution than the default. This dataset is built using data from multiple sources, resolution and geographical projections, as described below. We first project (to the nearest neighbor) all surface specification variables of interest to a common grid of a much higher resolution than the desired SURFEX grid on which we then compute spatial averages for each 400 m square grid cell. Table 1 lists the average values inside the city of Philadelphia of the different parameters we updated compared to what ECOCLIMAP would have given.

2.2.1. Surface type fractions of different tiles

The first step is to improve the representation of the fractions of the different tiles: town, nature, inland water, and sea and ocean. To do so, we use the 2019 version of the National Land Cover Database (NLCD) “Urban Imperviousness” product of the Multi-Resolution Land Characteristics Consortium (MRLC) (Dewitz and U.S. Geological Survey, 2021; Xian et al., 2011; Yang et al., 2018; Table 2) to derive the fraction of town (Fig. 2). The urban imperviousness dataset has been produced using a combination of Landsat 7 ETM+ images (30 m), IKONOS images (1 m) and a regression tree. For details, see Yang et al. (2018). The water fraction is defined using the Global Water Surface Explorer dataset (Pekel et al., 2016) and the fraction of nature is the residual between town and water. As the datasets we use were produced using different methodologies and at different times, it is possible that some points conflict (i.e. sum of fractions greater than 1). In this case, water takes precedence over urban imperviousness and the natural fraction is calculated thereafter as the residual. One more step is necessary to represent the appropriate amount of town: a portion of the natural areas needs to be put inside the town as urban vegetation called “garden” (Fig. 2).

2.2.2. In-town fractions and parameters

Inside the town part of every grid cell, SURFEX requires information about the fraction of buildings, the fraction of gardens and the residual component which is the fraction of roads. The fraction of buildings is derived from Heris et al. (2020)’s dataset available over

Table 1
Average (and 90 % range of grid cell values) of the updated surface parameters compared to the one provided by ECOCLIMAP inside Philadelphia city limits. Fractions are expressed over a complete grid mesh. Note that ECOCLIMAP has a single value for every parameter and that the ranges are due to the presence of some water and nature fractions inside the city.

Variable / parameter	ECOCLIMAP I v1.9	Updated dataset
Town fraction	0.91 [0.33–1.00]	0.77 [0.10–1.00]
Nature fraction	0.04 [0.00–0.37]	0.19 [0.00–0.88]
Water fraction	0.04 [0.00–0.38]	0.05 [0.00–0.35]
Building fraction	0.28 [0.17–0.30]	0.20 [0.01–0.42]
Road fraction	0.28 [0.17–0.30]	0.39 [0.07–0.94]
Garden fraction	0.38 [0.23–0.40]	0.21 [0.00–0.49]
High vegetation fraction	–	0.03 [0.00–0.48]
Low vegetation fraction	–	0.15 [0.00–0.31]
Bare soil fraction	–	0.04 [0.00–0.08]
Tree height (m)	–	4.01 [0.00–11.75]
Building height (m)	10.00	8.51 [5.24–15.09]
Wall area over horizontal plane area (m^2/m^2)	0.30	0.08 [0.03–0.20]

Table 2
List of all the datasets used to derive the surface boundary conditions necessary for SURFEX.

Dataset	Resolution	Reference	Used to define
ECOCLIMAP I	1 km	Champeaux et al., 2005	Physical parameters of natural and artificial areas
NLCD Urban Imperviousness	30 m	Dewitz and U.S. Geological Survey, 2021; Xian et al., 2011; Yang et al., 2018	Town fraction
Global Water Surface Explorer	30 m	Pekel et al., 2016	Water fraction
Heris et al., 2020	30 m	Heris et al., 2020	Building fraction
City of Philadelphia's Building Footprints	–	City of Philadelphia, 2015	Building fraction and building
NLCD Tree Canopy Cover	30 m	Coulston et al., 2012	High (tree), low and no vegetation fractions
Global Forest Canopy Height	30 m	Potapov et al., 2021	High vegetation (tree) height
Falcone, 2016	30 m	Falcone, 2016 and Demuzere et al., 2020	Building height

the whole CONUS. Inside the city limits, it is updated using the higher resolution City of Philadelphia's Building Footprints' dataset (2015). Since there is no obvious way to make the distinction between vegetated areas that are "natural" (e.g. forests) from the ones that are "artificial" (e.g. lawns, isolated trees, some urban parks) from the residual fraction, we follow Bernard et al. (2022)'s approach: if the fraction of buildings is greater than 2 % of the town then all the vegetation is moved into the garden; otherwise, only twice the fraction of buildings is converted. The fraction of town and nature are then updated accordingly. The fraction of roads is subsequently defined as the residual of buildings and gardens.

Inside the garden, three types of urban vegetation are defined: high vegetation (trees), low vegetation (grass and shrubs) and no vegetation (non-impervious bare soil). The proportion of trees is taken from the 2016 version of the U.S. Forest Service tree canopy cover product (NLCD, MRLC, Coulson et al. 2012) and first converted to a fraction of high vegetation over total amount of vegetation in the grid mesh (the nature fraction) without considering if it belongs in the garden or not. For each point, 80 % of the remaining vegetation is put as low vegetation and the other 20 % as bare soil. The last step regarding urban vegetation is to define the average tree height. This is done using the Global Forest Canopy Height, 2019 database (Potapov et al., 2021).

The last parameter needed to define the urban canyon geometry is the height of buildings, which is used to derive the town surface roughness length and the wall area over horizontal plane area. Outside of the city limits we use Falcone (2016)'s dataset that classifies buildings in the U.S. into 6 categories based on their estimated height derived from the NASA Shuttle Radar Topography Mission at 30 m resolution. We translate these classes to average height using Demuzere et al. (2020) estimates rounded to the nearest 3 m height, and set the lowest category to 3 m; ending up with: 3, 12, 12, 15, 21, 36 m average building heights (original values: 11.5, 13.1, 16.3, 21.7 and 35.3). Inside the city's limits, the same 2015 Building Footprints dataset that improved the building fractions is used to update the building heights.

Other variables, such as types of vegetation, physical parameters of vegetation (e.g. albedo, emissivity, LAI), are provided by ECOCLIMAP, with some of which evolve according to seasonal climatologies.

The updated dataset reduces the fraction of town inside the limits of the city (Table 1) compared to ECOCLIMAP (77 % vs 91 %) because urban parks, such as Wissahickon Valley Park, Fairmont Park, and Pennypack Park, are partially kept as nature instead of garden given their important sizes (which would otherwise result in unrealistic urban canyons with too much garden area, and consequently unrealistic geometries, i.e. too wide). The average fraction of buildings is also reduced in favor of a better representation of roads, sidewalks and parking spots. Perhaps the most important improvement is the range of values represented: with greater maximum urban fractions found in the high-rise dense city center. We note the use of two different datasets create a discontinuity in and outside of the city limits, which is most visible for average building height on the eastern side of the Delaware river in New Jersey.

2.3. *Atmospheric forcing*

As described previously, SURFEX is here run uncoupled to an atmospheric model, i.e. with the atmospheric conditions prescribed above the Urban Canopy Layer. In this configuration SURFEX can be seen as a sum of separate single columns (one for each point) that do not exchange fluxes between one another. The required atmospheric forcing variables are: air temperature, specific humidity, air pressure, rain and snow rates, wind speed and wind direction, downwelling long wave radiation, and downwelling direct and scattered short wave radiation. They can be prescribed at different temporal resolutions and SURFEX interpolates them to the model resolution.

For this study, the prescribed historical period (1991–2020) conditions are derived from the ECMWF global reanalysis ERA5 (Hersbach et al., 2020). A spin up period of two years (1989–1990) is chosen to initialize the model. The ERA5 reanalysis is based on the Integrated Forecasting System (IFS) Cy41r2 and covers the globe with a horizontal resolution of 31 km and a temporal resolution of 1 h. ERA5 has been used previously to drive SURFEX for vegetation (Albergel et al., 2018; Nogueira et al., 2020a, 2020b), hydrological (Albergel et al., 2018) and urban (Nogueira et al., 2022) applications. In addition to outputting surface variables, ERA5 also provides different atmospheric levels that we use to vertically interpolate temperature, humidity and pressure to 100 m. Every variable is also horizontally interpolated from 31 km to the SURFEX grid (400 m) using a bilinear interpolation. Table 3 shows that ERA5 reproduces the average daily summer minimum and maximum temperature extremely well at the location of Philadelphia International Airport's weather station (PHL; Fig. 1.b), with only a slight underestimation (max 0.6 °C in August for daily maximum temperature). This small

underestimation is most likely due to a combination of the absence of artificial covers and the resolution of the ERA5 product compared to smaller-scale variations in the area that surrounds the PHL weather station.

3. Evaluation

3.1. Evaluation against long-term reference observations

Ideally, one would like to evaluate urban climate simulations via comparisons with a collection of long-term measurements of the urban environment (Landsberg, 1970). In practice, such observations are often lacking because of the heterogeneity of the urban landscape and the difficulty to set up stations in representative areas (Oke, 2004). Depending on the size of the city studied and the aim of the work carried out, the closest airport is often used as the reference urban station, with caveats highlighted that doing so often misses parts of the UHI. For Philadelphia that station is located at the Philadelphia International Airport (PHL), approximately 11 km southwest of the city center and Philadelphia City Hall. It is part of the NOAA Integrated Surface Database (ISD) providing hourly measurements of near surface (2 m) air temperature, dew point and other variables from 1940 to present. Two other stations that are part of NOAA Regional Climate Centers Applied Climate Information System (ACIS) are used to compare daily minimum and maximum values of temperature and humidity: the station of the Northeast Philadelphia Airport (PNE; ~19 km northeast of city center, with less than 1 % missing over the period 1991–2020), and the rooftop station of the Franklin Institute (PFI) located in the city center (data available since August 1993 with less than 3 % missing).

Fig. 3 represents the average daily cycle of summer (JJA) near surface air temperature observed at the Philadelphia International Airport and simulated by SURFEX for the points closest to the three weather stations. For the PHL station location (Fig. 3.a), SURFEX is able to represent the daily cycle of temperatures quite well, without any systematic bias when averaged over the 24 h: the model is only 0.1 °C too cold (24.6 °C against 24.7 °C). However, when distinguishing between different periods of the day, we see that the model underestimates nighttime temperatures and overestimates them during the day. The maximum biases are found at 6:00 EST with SURFEX being 1.1 °C too cold (19.9 °C vs 20.8 °C) and then at noon with +1.4 °C (29.1 °C vs 27.7 °C). Though the daily minimum is captured at the correct hour (05:00 EST), the simulation warms too quickly during the morning and reaches its maximum one hour too early (15,00 EST).

The differences between modeled and observed temperatures at PHL can be due to some systematic SURFEX biases, but can also arise because the 400 m SURFEX model grid cell cannot perfectly represent the station's particular geographic situation. Indeed, the PHL station's instruments are located a few dozen meters away from a paved runway, on an open surface of a mix of dirt and gravel, while the corresponding SURFEX point is composed of 46 % of town and 54 % of nature. No vegetation is represented inside the fraction of town, instead almost everything is described as road – 98 % – with the remaining 2 % representing buildings. This high amount of impervious surfaces represent the multiple runways, taxiways, and other road-paths that surround the station over a 400 m grid cell and likely contribute to the positive bias observed during the day. With the weather instruments located directly above a mix of dirt and gravel, and grass areas between the instruments and airport runways and taxiways, the station is expected to warm less under sunny conditions compared to the roads in SURFEX.

No hourly measurements are available for the PFI station, but for illustrative purposes Fig. 3.b and c represents what SURFEX simulates for the corresponding points. The PHL and PNE stations are located 10 km and 19 km away from the Franklin Institute, to the south-west and north-east, respectively. In addition to the 2 m air temperature, Fig. 3.b also shows what is simulated at 25 m inside the urban canyon, which corresponds to the height of the rooftop station above the street. Averaged over the summer months, the PFI point is warmer than both airports at all hours, especially at night and near the surface (2 m). Temperatures are lower at PNE than at PHL by ~0.5 °C during the day. Not taking into account possible discrepancies in the forcing, in terms of surface type fractions the two airport grid cells are almost identical, with the latter having 45 % of town, 55 % of nature and 98 % of building inside the town. The only notable difference between the two airport grid cells is the ratio of wall surfaces over horizontal plane area that defines the canyon

Table 3
Average (and 90 % range) values of summer daily minimum and maximum temperatures, daily temperature range and precipitation observed at the Philadelphia International Airport and found in the closest ERA5 point.

		ERA5			Philadelphia International Airport station (PHL)		
		Avg	p05	p95	Avg	p05	p95
TN (°C)	June	18.4	12.3	23.6	18.5	12.8	23.3
	July	21.3	16.4	25.5	21.6	17.2	25.6
	August	20.4	15.3	24.7	20.8	16.1	24.4
TX (°C)	June	27.8	21.2	33.6	28	21.1	34.4
	July	30.1	25.2	34.7	30.6	24.4	36.1
	August	29	24.4	33.7	29.4	23.9	34.4
dtr (°C)	June	9.4	4.7	14	9.5	4.5	13.9
	July	8.8	5	12.4	9	5	12.3
	August	8.6	4.4	12.6	8.7	4.4	12.2
pr (mm per month)	June	3.2	0	16.9	2.8	0	16.9
	July	3.0	0	16.5	3.1	0	15.0
	August	3.2	0	16.7	3.0	0	19.9

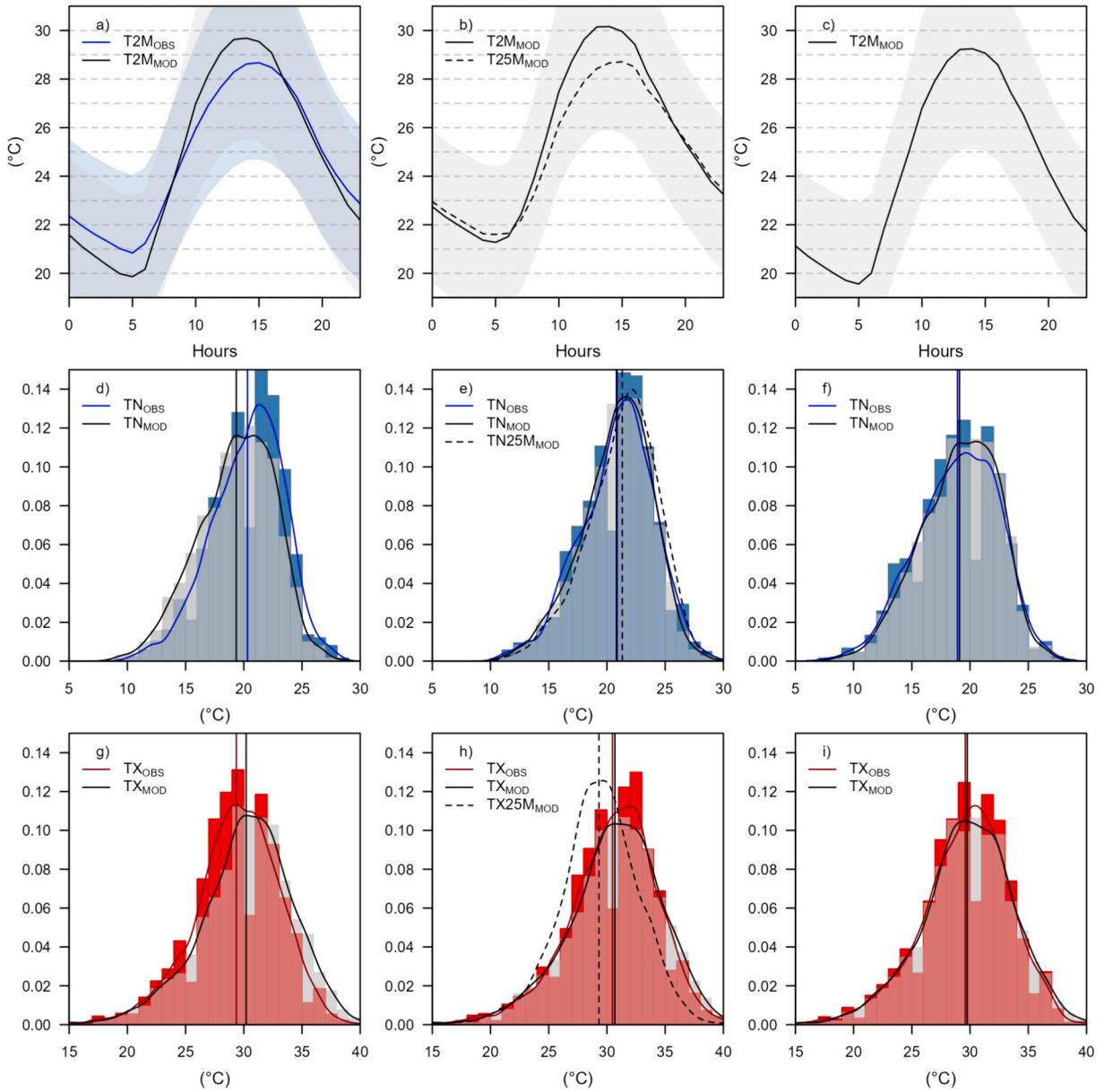


Fig. 3. Comparison of the average daily cycles and daily minimum and maximum distributions of summer temperatures observed at the three stations and modeled at the closest SURFEX grid point (Philadelphia International Airport [a, d, g], Franklin Institute [b, e, h] and Philadelphia Northeast Airport [c, f, i]). The grey areas in panels a, b and c represent one standard deviation above and under the mean. The vertical lines in panels d to i represent the averages of the distributions. The model distributions in panels e and h are for TN and TX at 2 m, the dashed lines show the distributions for TN and TX at 25 m. Note that for panels d to i, the "dips" in the observation histograms visible at 5 °C intervals are due to the observations being recorded in integer degrees Fahrenheit, rounded to one decimal place Celsius, and plotted here using 1 °C wide bins.

geometry: 0.44 for PHL vs. 0.11 for PNE. The relatively important difference between the two is because the former is associated with a higher building height that was determined using the lower resolution dataset of Falcone (2016), compared to the updated database for PNE.

Fig. 3.d through i represents the seasonal distributions of daily minimum (TN) and maximum (TX) temperatures simulated by the model and compared to observations for JJA. Note that for the Franklin Institute (Fig. 3.e and h) the distributions of TN_{25m} and TX_{25m} are also displayed (dashed lines). Given that the timing of the maximum temperature between the model and the observation is slightly shifted, it is expected that the bias in TX will be smaller than the 15:00 EST bias. Indeed, for the PHL station the average biases are −1 °C (19.4 °C vs 20.3 °C) for TN and 0.8 °C (30.2 °C vs 29.4 °C) for TX. The biases get slightly exacerbated on the left and right tails of the distribution for TN and TX respectively, i.e. the model underestimates the coldest minima and overestimates the warmest maxima

(P_{10}TN bias = −1.2 °C with 14.8 °C vs 16.1 °C and P_{90}TX bias = 1.2 °C with 35.1 °C vs 33.9 °C). Nevertheless, even at those extremes the biases remain small for the PHL station and overall SURFEX simulates the shapes of the distributions extremely well (Skewness (Geary, 1936): TN_{MOD} = −0.37 vs TN_{OBS} = −0.42; TX_{MOD} = −0.49 vs TX_{OBS} = −0.48; Kurtosis (Pearson, 1905): TN_{MOD} = 2.8 vs TN_{OBS} = 3.1; TX_{MOD} = 3.6 vs TX_{OBS} = 3.5). The results at the PNE airport in the northeast are even closer to observations with average biases of 0.2 °C for both TN (19.1 °C vs 18.9 °C) and TX (29.8 °C vs 29.6 °C).

At the Franklin Institute, at night the model is on average only 0.1 °C colder than the observations (20.8 °C vs 20.9 °C). The better results observed at this site in the city center could be due either to a better correspondence between the SURFEX point characteristics and the area surrounding the institute or to the fraction of the point that is considered as town, which would then hint that the biases could be coming from the ISBA vegetation module rather than TEB. Given that the PFI is located on the rooftop ~25 m above ground, the temperatures at that model level are also displayed. Because the air is well mixed inside the canyon at night, both levels show similar TN values (Fig. 3.e). During the day, more differences are noted between the two levels and consequently with the observations

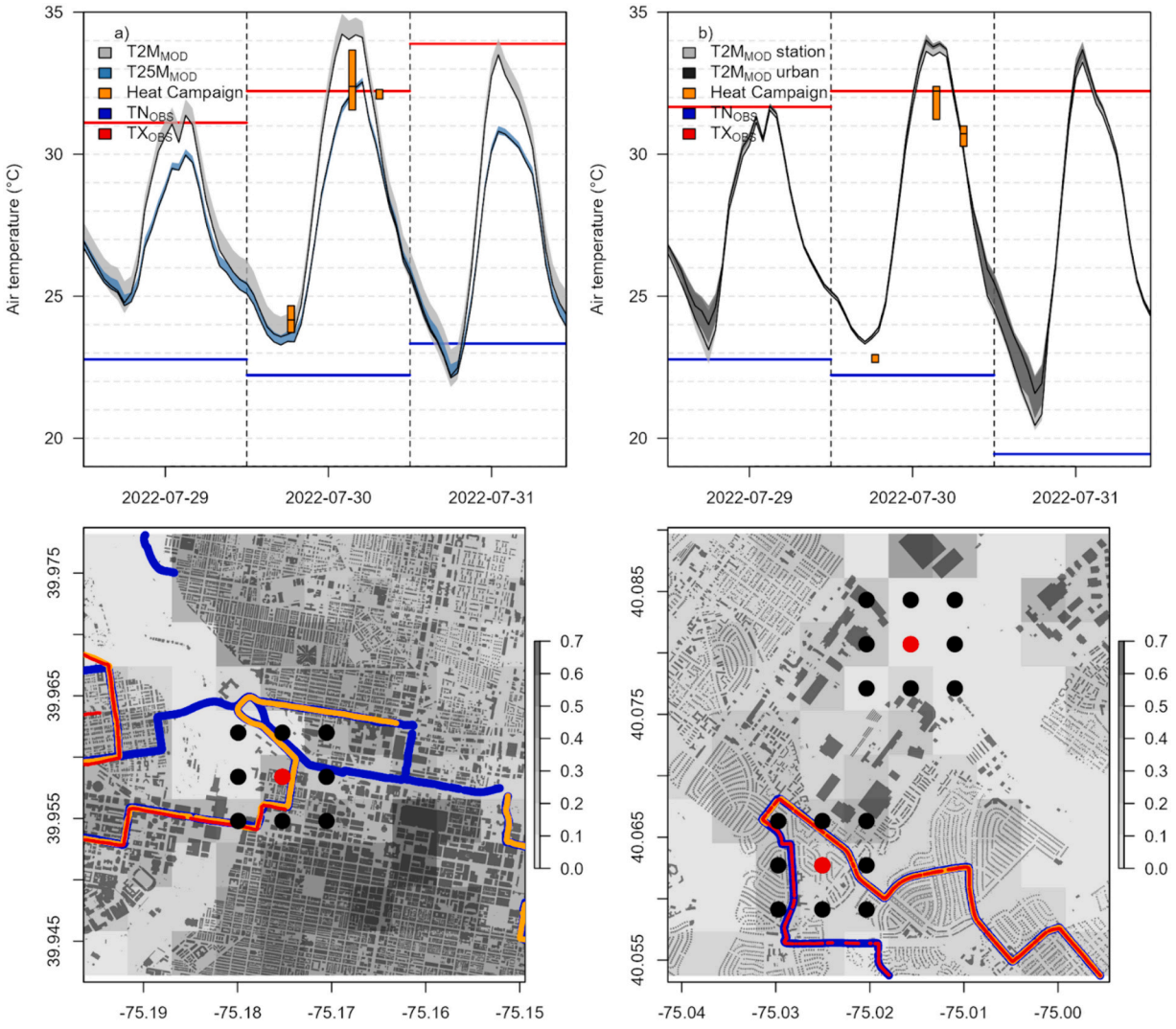


Fig. 4. Comparison of hourly near surface temperatures simulated by SURFEX (curves, top row) near the PFI (left column) and PNE stations (right column). Red dots (bottom panels) indicate the SURFEX grid points closest to the two weather stations, and for PNE a red dot to the SW of the airport was selected for its proximity to the heat campaign track. For the PFI station, the SURFEX temperature at 25 m is shown as a curve. Grey areas surrounding the SURFEX curves represent the range of temperatures simulated on the 8 grid points (black dots) surrounding their respective red dots, excluding points in water. The stations' observed daily minimum and maximum temperatures appear as horizontal lines (top row). Color-filled rectangles depict temperatures recorded by the mobile heat campaign within 300 m of their respective weather stations. The heat campaign tracks are shown as color-coded paths in the bottom panels (blue, orange and red for morning, afternoon and evening, respectively). The 3 colored paths are drawn at different widths so that, when the same segment was covered multiple times, earlier paths appears as colored lines at the edges of the later path. The grey scale represents the fraction of buildings in SURFEX. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fig. 3.h). $TX_{25m,MOD}$ is −1.2 °C smaller than TX_{OBS} (29.3 °C vs 30.5 °C) while the TX_{MOD} at 2 m is only 0.2 °C warmer than TX_{OBS} (30.7 °C vs 30.5 °C). Even though $TX_{25m,MOD}$ is assumed to be at a height closer to observation, the differences are most likely due to the radiative effect of nearby surfaces not captured by the model simulating the temperature in the middle of the canyon street (reflection and heat emission from the roof).

3.2. Comparison to a short term field campaign

The difficulty of comparing high resolution climate simulations to observations in urban environments was highlighted earlier: e.g.

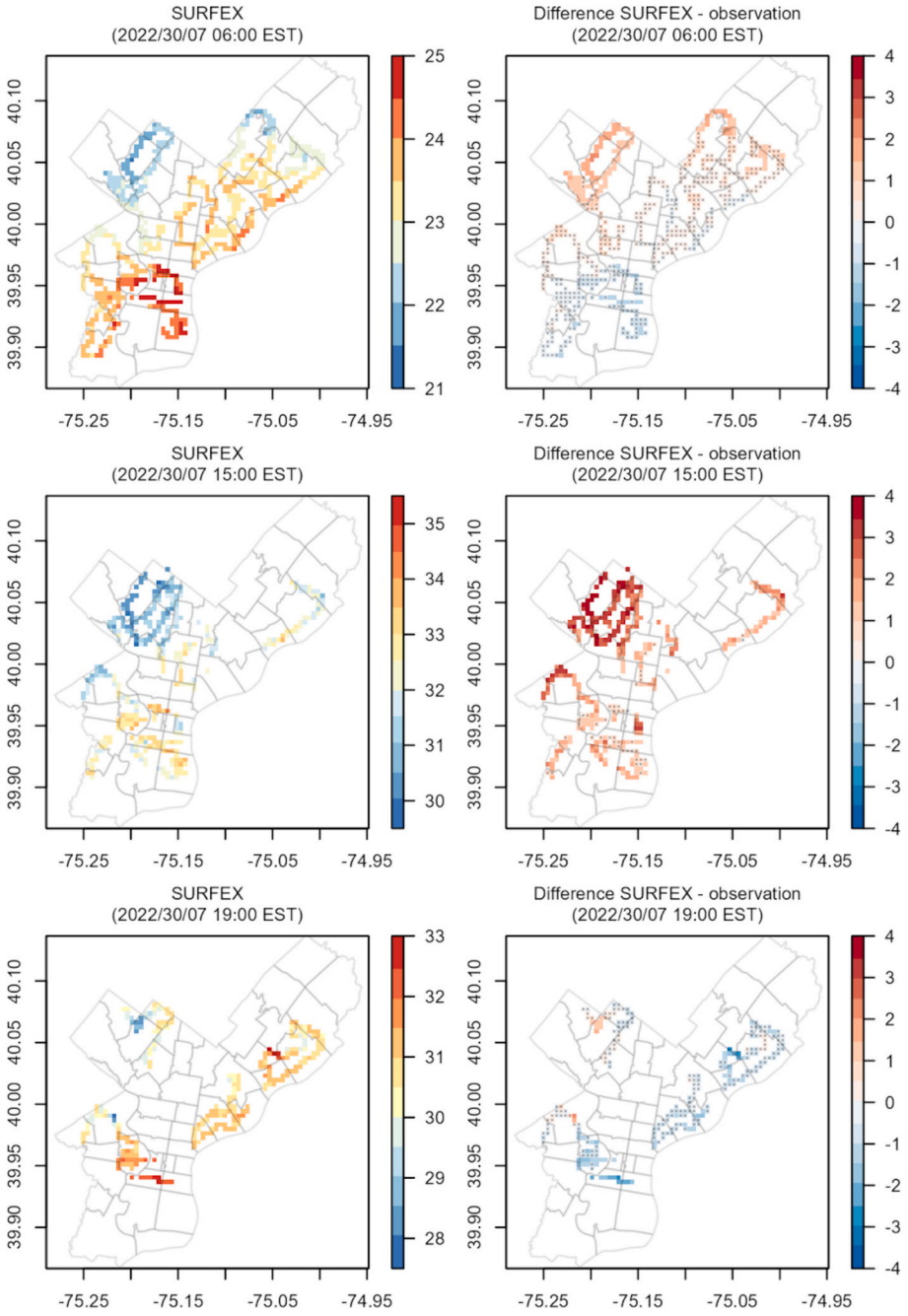


Fig. 5. Comparison of the near surface temperature simulated by SURFEX (left column) on July 7th 2022 in the morning, afternoon and evening and differences (right column) with NOAA's Heat Campaign measurements. The black dots represent model points with an absolute bias smaller than 1 °C.

representativity of one point measurement to the heterogeneity of a building block. While one well-chosen reference rural station potentially could be considered sufficient to represent the background climate surrounding the city, another issue is the more general lack of urban stations located in multiple parts of the city with different characteristics. One approach sometimes chosen to address this issue is the use of remote sensed measurements - usually coming from satellites - which has its own limitations, including temporal frequency and representativity of the urban facets depending on the resolution and the viewing angle of the satellite. Also, the accuracy of inferring 2 m air temperatures from satellite retrievals can be hindered by the complex heterogeneity of the urban environment at small spatial scales (Yoo et al., 2018).

Here to qualitatively examine the ability of SURFEX to simulate different parts of the city, temperatures recorded during the 2022 Urban Heat Watch mapping campaign in Philadelphia are used (CAPA Strategies, 2022; Shandas et al., 2019). The campaign was carried out the 30th of July 2022, with volunteers driving around 3 transects in the city for around one hour, one in the morning (06 EST), one in the afternoon (15 EST) and one in the evening (19 EST). To allow comparison with model results, the SURFEX simulation was extended from 2020 to that day.

Fig. 4 presents, for 30 July 2022, a comparison of three sources of temperature information; namely, daily minimum and maximum station observations at the PFI and PNE stations (horizontal lines), hourly SURFEX model results near those two sites (black curves), and nearby data gathered during the mobile heat campaign (color-filled rectangles). The routes driven during the mapping campaign did not come within 3 km of the PHL station, precluding comparisons of the type depicted in Fig. 4 for that location.

Considering the Franklin Institute area (Fig. 3a), at 6:00 EST SURFEX simulated a 2 m temperature between 23.7 °C and 24.4 °C, only 0.3 °C lower than the upper value recorded by the mobile transect (min-max: 23.7 °C - 24.7 °C); while the minimum temperature observed the night before at the PFI station was 22.8 °C. During the day, the mobile transect shows a lot of variability in the temperatures observed around the station (between 31.6 °C and 33.7 °C) in agreement with the maximum at the station (32.2 °C) and the 25 m temperature simulated by SURFEX (32 °C). However, SURFEX overestimates the 2 m temperature by 1.6 °C at 15 EST. Lastly, while SURFEX simulates a steep decrease of temperature after 18 EST, the measurements made in the evening (19 EST) are quite close to the central value of the afternoon, with much less spatial variation.

Fig. 4.b represents the same comparison, but for the PNE station and nearby transects. Because the PNE weather station is located further away from the transect than is the case for the PFI station, two sets of 9 points are compared ($T2M_{MOD}$ station and $T2M_{MOD}$ urban). During the day, the two sets of PNE points exhibit almost identical modeled air temperature and at night the station points can be colder by around 1 °C maximum. Similar results to the city center are found: SURFEX does better during the night (with nevertheless a slight overestimation of the morning temperature here) than during the day.

Thanks to their spatial coverage, the heat campaign measurements make it possible to evaluate if the biases observed near the reference stations are also present elsewhere in the city. For each transect hour, they are interpolated to the SURFEX grid by averaging all the observations inside a grid cell. The interpolated maps are then compared to the average SURFEX maps of the hours around the campaign (e.g. 06:00 EST and 07:00 EST for the morning). Fig. 5 represents those SURFEX maps as well as the difference between the model and the mobile heat campaign observations. The model represents spatial variations around the city, with the denser city part in the south being the warmest just before sunrise (up to 25 °C), and the more vegetated part in the northwest (Wissahickon Valley Park) being colder by around 3 °C. Similar spatial differences are observed in the afternoon and the evening, with slightly smaller relative differences.

Compared to the observations, in the morning the model bias is inversely correlated to the building density. The densest part of the city is on average cooler in the model than in the observations - and the absolute biases are smaller - while the northernmost part is warmer, with stronger biases. Still, with the exception of the parks in the north, absolute biases are on average smaller than 1 °C. In the afternoon the overestimation of the temperature by SURFEX noted previously around the Franklin Institute is also found across the city, with densely built areas having biases around 1 °C and parks up to 4 °C. As noted previously, in the evening the model generally cools more quickly than the observations and consequently slightly underestimates the temperature; nevertheless, the biases are small (less than 1 °C).

The evaluations presented in sections 3.1 and 3.2 show that the SURFEX simulation is able to capture the diurnal cycle of summer temperatures well, with a modest underestimation of nighttime temperatures and overestimation of daytime temperatures. Observed differences between the airports and the city center's daily minimum and maximum temperatures are also well simulated by the model. Temperatures in more densely urbanized areas appear to be better simulated, especially at night when almost no bias is found. Somewhat greater biases are found during the day over less densely built parts of the city potentially highlighting the effect of the vegetation. Overall, the results are extremely satisfactory and in line with previous work carried out over Philadelphia (Ramamurthy and Bou-Zeid, 2017).

4. Discussion

4.1. The heat index as a measure of thermal comfort

To study the thermal comfort of inhabitants, a version of the Heat Index (HI) originally defined by Steadman (1979) is used. The HI represents an apparent temperature computed from the dry air temperature (T2M) and the relative humidity (HU2M). Steadman (1979) first developed a physical model to compute the index based on meteorological conditions, clothing and skin thermal resistance of a person; thereby creating a table displaying the expected HI for a range of temperatures and humidities. Since then, different formulations have been proposed to compute HI values (e.g., Anderson et al., 2013; Lu and Romps, 2022). Here, the heat.index function of the weathermetrics (Anderson et al., 2016) R package (R Core Team, 2022) is chosen, which mimics the procedure adopted

for operational use by the US National Weather Service (NWS), which in turn is based on the polynomial fit developed by Rothfusz and Headquarters (1990), with some modifications.

For this evaluation, the daily maximum HI is the principal diagnostic measure of interest. HI values are computed from 24 hourly temperature and humidity values and then the daily maximum is determined for each day. Attention is given to simulated and observed HI values for the PHL station/point for two reasons; namely, because PHL records hourly temperature and humidity measurements and also because the Philadelphia Health Department bases its Heat and Health Emergency declarations on publicly available HI forecasts for the airport issued by the NWS (Philadelphia Dept. of Health, personal communications).

Because of the temperature biases noted in the previous section, it is expected that the HI will also be biased to an extent. Fig. 6 shows the average summer daily cycle of T2M, HU2M and HI observed at the PHL weather station and simulated by SURFEX at the closest point to the station location, as well as the distributions of T2M and HU2M when HI is at its daily maximum value.

On average, during JJA over the 1991–2020 period, SURFEX simulates HIX_{JJA} values that are 0.6 °C too warm (31.0 °C vs 30.4 °C). The biases are similar over most of the distribution, with the most extreme HI values being simulated well (the 99th percentile has a bias smaller than 0.1 °C). The HI biases are smaller than the ones noted for the maximum temperatures because the maximum HI doesn't always necessarily happen when the temperature bias is the greatest during the day (grey areas in Fig. 6). Also, the SURFEX mean relative humidity is slightly too dry during the day (46.4 % vs 49.3 %). We note that HI variations are more affected by changes in temperature than humidity, which is well captured by the model.

Though the overestimation of HIX is relatively small, it can nevertheless have an impact if it is used to define different heat events on the basis of threshold exceedance. To address the SURFEX simulated HI bias issue, a univariate bias correction method is applied directly on HIX based on the quantile delta mapping QDM algorithm of Cannon et al. (2015), with the HIX computed at the airport serving as observational training data (more details on QDM in Lanzante et al., 2020). Fig. 6.f compares the distributions of HIX_{JJA} simulated by SURFEX before and after adjustment to the observations. As expected from a quantile-based correction method, after adjustment, the distributions are almost identical (remaining differences are smaller than 0.02 °C on average over all percentiles).

4.2. Philadelphia's heat caution and heat emergency declarations

Since 2017, the Health Department of the city of Philadelphia has been issuing "Heat Health Emergency" (HE) warnings during the extended summer season (May to September). The HE warnings are generally issued a few days in advance, and the objective part of

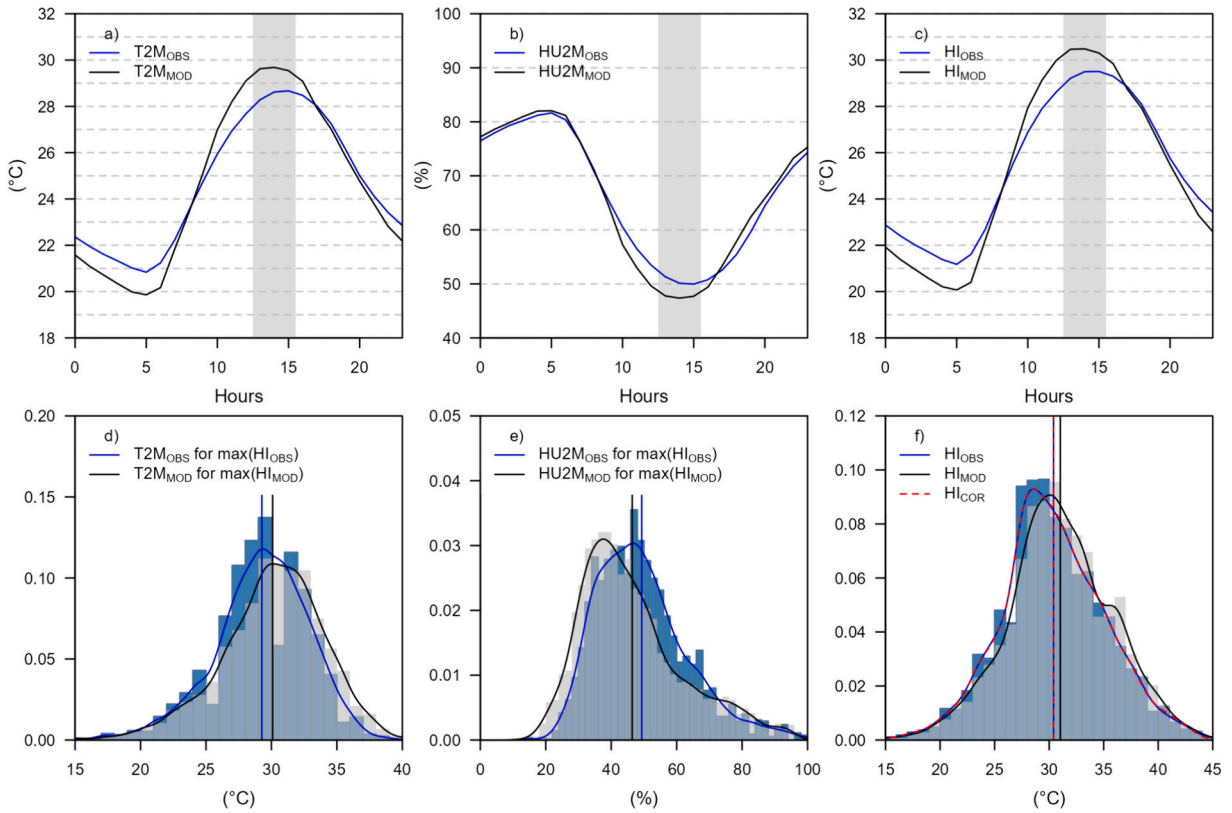


Fig. 6. Observed and simulated summer (JJA) averages diurnal cycles over the period 1991–2020: (a) 2-m temperature, (b) 2-m relative humidity and (c) corresponding Heat Index. Panels d, e and f represent the distributions of the corresponding variables at the hour where the Heat Index is maximum. Dips every 5 °C in the panel d histogram arise due to conversions of observations recorded as integer °F, as described in Fig. 3.

the process is based on the public forecast made by the NWS for Philadelphia International Airport. (In borderline situations, officials may also consider subjective, non-meteorological factors when deciding whether to issue a HE Warning.) Following objective criteria, for the city to declare an emergency the HI forecast must be above the different thresholds presented in Table 4, depending on the time of the year and the expected duration of the heat event. Lower thresholds are chosen for longer events to take into account the possible adverse health effects of the accumulation of heat stress over multiple days, while higher thresholds are defined for shorter events and for the end of the summer to take into account possible acclimation during the season. The HI thresholds were determined by a retrospective analysis conducted by the Philadelphia Department of Health that examined the relationship between the number of people going to hospitals in the city due to heat related illnesses and the observed HIX measured at the airport (Philadelphia Dept. of Health, personal communications).

In our analyses, we adopt the same HIX thresholds used operationally by the city. However, we expect to find slight differences between days we define as Heat Caution and Heat Health Emergency vs. days the city made official declaration because the city bases its declarations on NWS HI forecasts whereas we do a retrospective analysis of the airport observations. Additionally, in practice city officials can introduce subjective judgements, such as choosing to keep an event's warning going if, for example, one day in the midst of a heat wave has a forecast HIX value that falls just slightly under a threshold but the forecast projects that the event will continue (exceed HIX thresholds) on following days. Note that while we display values in degrees Celsius herein, the thresholds are defined, and consequently our computations are made in degrees Fahrenheit.

Based on airport weather records over the 1991–2020 period, the criteria for Heat Caution events (HC) were met 67 times, of which 13 exceeded the criteria for Heat Health Emergency (HE) events (Table 5). As expected, on average HC last longer than HE (3.5 days vs 2.7 days) and reach lower maximum HIX (40.5 °C vs 44.2 °C). All of these events occurred during the five-month extended summer season (May through September), and most of them in July and August. Panels a and b in Fig. 7 represent "bubble charts" comparing the duration, intensity (maximum HIX value reached during the event), and severity (defined as the sum of degrees above the detection threshold) of every HC event (including HE) calculated from three time series; namely, from the PHL weather observations and from temperatures and humidity simulated by SURFEX for the PHL grid point with and without bias correction.

Before bias correction, SURFEX simulates too many HC (91 vs 67) that last too long on average (3.7 days vs 3.5 days; note that this average duration is influenced by the overestimation of short 2 and 3-days events simulated by SURFEX). Consequently, the severity of the events is also slightly overestimated by the model (mean of 28.6 vs 27.8). As noted previously, the model biases are relatively small and do not markedly impact the maximum intensity and severity of the events. In fact, the highest HIX intensities are better simulated than the average and consequently we find that SURFEX slightly underestimates the average maximum intensity of the events, by 0.4 °C on average. Similar results are found when looking at HE only, with even smaller biases.

Bias-adjustment of HIX leads to fewer HC events being simulated (79 rather than 91) and fewer HE events (11 rather than 14). Days that were clearly defined as HC (i.e. for which HIX was clearly above the detection thresholds) are not impacted by the bias adjustment. Other events with days that were barely above the thresholds see their duration reduced by one or two days (Fig. 7.c); this greatly reduces the number of 2 and 3-days events and also completely eliminates 8-days events (that can be found in the observations). Some events get split into two (points connected with the dotted lines in Fig. 7 panels c and d). When looking at timeseries (not shown here) almost all the events than can be observed are simulated by SURFEX. The model does miss some of the shortest duration ones and simulates some other ones that do not appear in the observations. For the majority of events, the bias adjustment increases the maximum intensity (as it was underestimated), e.g. for the most extreme event, SURFEX only simulates a HIX of 47.9 °C compared to 53.5 °C in the observations (Table 5); which is completely corrected after the bias adjustment. The events considered as most severe (wider points in Fig. 7) do not necessarily correspond to HE, because our definition of severity simply looks at the sum of degrees above the detection threshold and does not give more weight to higher intensities (i.e. a 4-day event 1 °C above threshold has the same severity as a 2-days event 2 °C above threshold).

Beyond the study of the heat conditions observed and simulated at the PHL airport site, the SURFEX simulations allow us to look at the heat stress conditions across all neighborhoods of the city during these particular events.

Fig. 8.a represents maps of the SURFEX-simulated average HIX conditions during the extended summer season over the period 1991–2020. (Note: because of the lack of spatialized urban observations or reanalysis we can not apply the QDM correction to every SURFEX point. The urban HIX maps therefore have not undergone bias correction processing.)

The highest HIX are found in the densest part of the city center, near the city hall. The HIX pattern resembles what one might expect for nighttime air temperature UHI, but less so for the HIX that is based on maximum air temperature. Maximum UHI intensities are

Table 4
Daily maximum Heat Index thresholds used by the Philadelphia Health Department to issue heat related warnings during the extended summer season (May to September).

		Short duration level		Long duration level	
Heat Caution	May 1 – June 30	2 days	≥ 95 °F ≥ *35 °C*	≥3 days	≥ 93 °F ≥ *33.9 °C*
	July 1 – September 30	2 days	≥ 98 °F ≥ *36.7 °C*	≥3 days	≥ 95 °F ≥ 35 °C
Heat Health Emergency	May 1 – June 30	2 days	≥ 101 °F ≥ *38.3 °C*	≥3 days	≥ 98 °F ≥ *36.7 °C*
	July 1 – September 30	2 days	≥ 106 °F ≥ *41.1 °C*	≥3 days	≥ 103 °F ≥ *39.4 °C*

Table 5
Characteristics of the different events observed at the Philadelphia International Airport and simulated by SURFEX (number of events over the period 1991–2020 May to September; average duration, average and maximum of the maximum HIX reached; severity defined as the sum of degrees above the detection thresholds).

		Number	Average duration (days)	Maximum Heat Index: average and (maximum) °C	Severity: average and (maximum) °C
Heat Caution	Observation	67	3.5	40.5 (53.5)	27.8 (85.3)
	SURFEX	91	3.7	40.1 (47.9)	28.6 (84.2)
	SURFEX ADJ	79	3.4	40.4 (53.5)	26.8 (81.5)
Heat Health Emergency	Observation	13	2.7	44.2 (53.5)	18.1 (38.6)
	SURFEX	14	2.7	43.0 (47.7)	18.3 (33.6)
	SURFEX ADJ	11	2.9	44.5 (48.6)	20.9 (35.8)

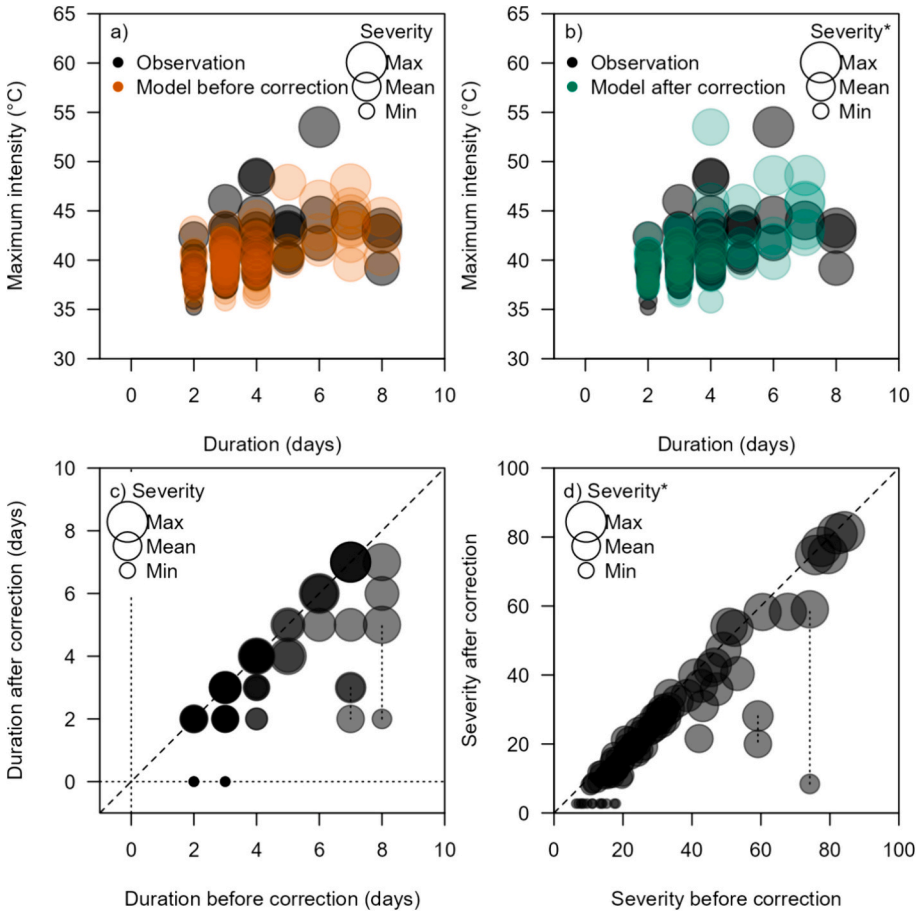


Fig. 7. A "*bubble chart*" of Heat Caution events (Heat Health Emergency also included) observed at the Philadelphia International Airport and simulated by SURFEX at the closest point. Bubble size represents event severity, which is the cumulative number of degrees Celsius by which an event exceeds the detection thresholds. Panels a and b compare the duration, maximum intensity and severity of each event, before and after bias correction. Panels c and d compare the effect of the bias correction on the durationand the severity of each event, respectively. The opacity of a point represents the number of events sharing the same characteristics.

usually expected at night rather than during the day when tall and densely built areas should be shaded from solar radiations and thus cooler than more open artificial surfaces. Yet we note the short term heat campaign measurements displayed in Fig. 5 suggests that the dense city center can indeed be warmer at the end of the afternoon. A few prior studies also have found positive daytime UHI in dense city centers (e.g., Soltani and Sharifi, 2017). The positive biases in model-simulated maximum temperature could be exacerbated in this specific area because SURFEX doesn't represent horizontal advection between grid cells, therefore points with tall buildings might accumulate too much heat during the day. Nevertheless, over the whole city (limits displayed in grey in Fig. 8), average HIX values are comparable to the value simulated at the airport (28.5 °C and 28.6 °C, respectively).

Besides the dense city center, other areas also experience higher than city-average HIX values in the SURFEX simulation. Inside the

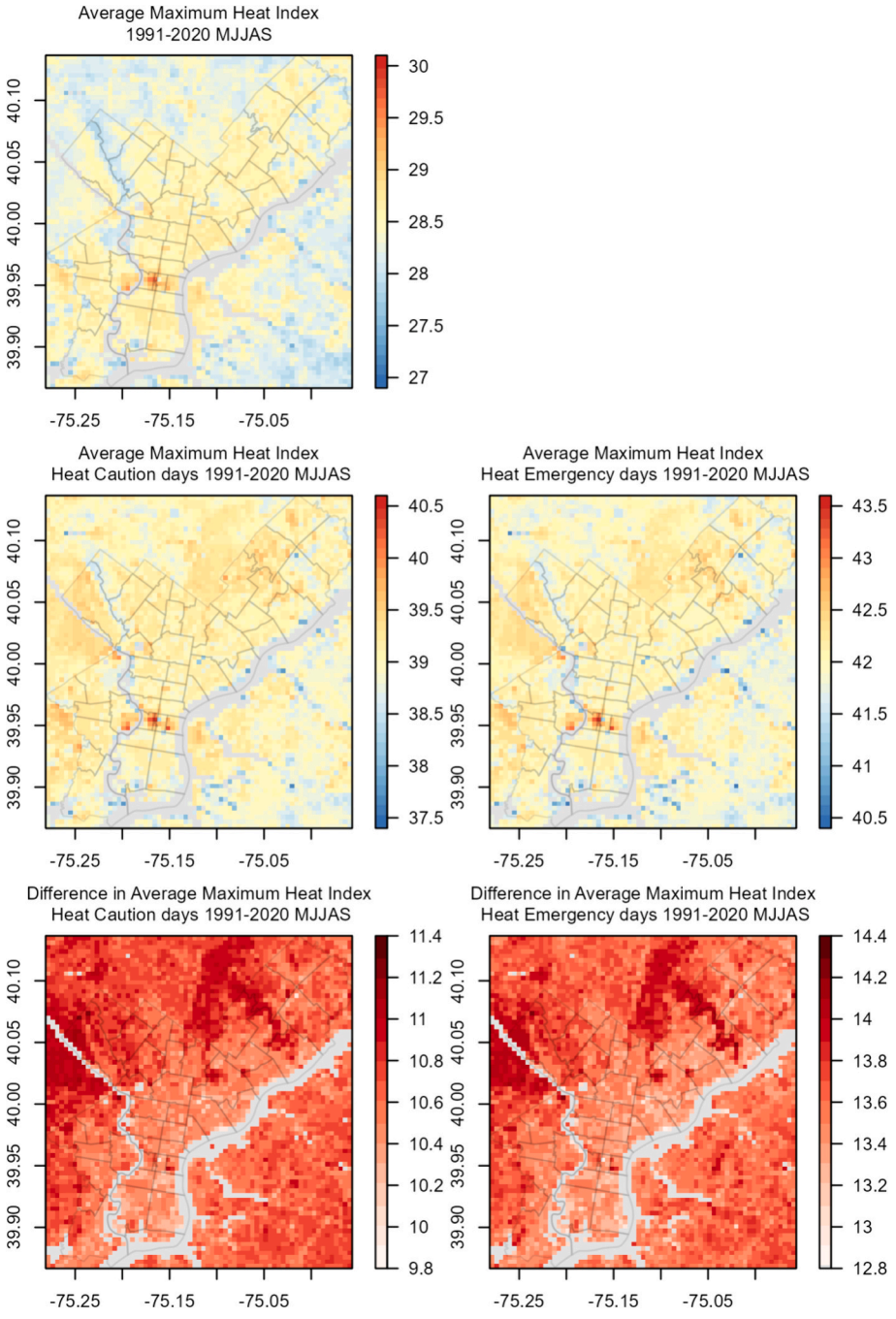


Fig. 8. Average map of daily maximum Heat Index simulated by SURFEX over the city of Philadelphia for the extended summer (May to September) of the 1991–2020 period, and for the days categorized as Heat Caution and Heat Health Emergency. The last row represents the differences between those particular days and the whole period average. Note that the color scales cover different ranges.

city, South Philadelphia is the most impacted and outside of Philadelphia's borders to the east, Camden, New Jersey has similar average heat stress conditions. Note that the differences in building datasets used inside and outside of Philadelphia's borders might explain some discrepancies. As it was observed in the previous section, areas with the highest vegetation coverage in the north-west exhibit the least UHI effects (mean HIX under 28 °C for the extended summer season).

Additional analyses restricted to model-simulated HC and HE days (based on the HIX simulated near the airport and bias adjusted) allows an examination of extremely hot days. During Heat Health Warning and Caution days, similar HIX values are simulated for the whole city and the point at the Philadelphia International Airport, with 39.1 °C (whole city) and 39.3 °C (airport) for HC days and

42.1 °C and 42.5 °C for HE days. Spatial patterns similar to those seen in the 1991–2020 average are also present both for HC and HE days, with the most impacted areas in the dense city center, in the south and to the east. However, the contrast between the whole city and less densely built areas is less apparent during HC and HE days than in the extended season averages. The only significant differences are found outside of the city centre (Fig. 8.d,e) in the most vegetated areas (Pennypack Park in the northeast and Fairmount Park to the west). Given that the HI is based on temperature and humidity, we should expect lower HI values where it is colder and dryer. When looking at humidity in the vegetated areas, we observe dryer conditions which should go in the direction of a decrease of the HI (all else being equal). Conversely we find that it is in fact higher than the rest of the area. This could be due to the fact that lower humidity values decrease the ability of the vegetation to cool the air through evapotranspiration, therefore increasing TX, and consequently HI because the contribution of temperature is greater than that of humidity to the index.

5. Conclusion

To study the evolution of the urban climate of Philadelphia, Pennsylvania, we developed a modeling framework making use of a high resolution database describing land surfaces and the urban geometry and a land surface model (SURFEX) incorporating a dedicated urban canopy model (TEB). The approach developed for Philadelphia could be adopted for any city in the country, because the surface boundary conditions are based on openly accessible datasets available for the whole continental United States, and the global atmospheric reanalysis ERA5 is used to drive the model. Flexibility in the model configuration process includes the ability to use city specific data products for some boundary conditions if they are deemed preferable to national data sets. However, we note that SURFEX simulations might not be suited for all cities depending on the geographical context, e.g. near coastlines where the absence of horizontal advection in the model will prohibit the simulations of sea breezes.

Model simulations were carried out over the 1991–2020 period and evaluated against long term observations for lower density portions of the city (near the Philadelphia International Airport) and inside the city center (over the Philadelphia Franklin Institute). Comparisons showed that the model simulated the daily cycle of temperature and humidity well (−0.1 °C over 24 h) with an underestimation of daily minimum temperatures (TN; −1.0 °C) and an overestimation of daily maximum (TX; 0.8 °C) at the PHL airport point. In the city center, results are better with almost no average biases (−0.1 °C for TN and 0.2 °C for TX).

To study heat stress conditions across the city, a retrospective analysis of the daily maximum Heat Index (HIX) was carried out. The HIX diagnostic measure (a function of temperature and humidity) was selected because it is used by the city of Philadelphia to declare its heat warnings during the summer season. Because of the biases highlighted previously, the model simulates more heat emergency events than what is calculated from airport observations, and the simulated events are on average slightly too intense (maximum intensity) and severe. Bias correction of SURFEX outputs was shown to improve agreement of simulated and observed heat events. When looking at spatial differences, while the average of the city experiences the very similar conditions as that simulated for the airport, the densest parts of the city in the center and the south are more impacted (i.e., higher HIX). The least impacted areas during average summer conditions are the ones with the most vegetation. During extreme heat events the whole city experiences worse conditions but without significant spatial changes; on the other hand, outside of the city, HIX values in parks worsen because of dryer conditions limiting the cooling effect of the vegetation through evapotranspiration.

This modeling study allowed us to show that when extreme heat stress conditions are measured at Philadelphia International Airport, they appear to be sufficiently representative that extreme conditions are being experienced across the city. This is consistent with the reliance of the Philadelphia Department of Health's approach to declaring citywide Heat Health Emergencies based upon heat index values for Philadelphia International Airport. Additionally, the model results potentially could be used as part of an effort to identify neighborhoods that experience higher HIX conditions during heat emergencies, and thus would benefit from specific attention during extreme heat events.

This study was a first step to develop and evaluate the ability of the SURFEX model to simulate the climate of Philadelphia in recent decades. Future work will use the same modeling approach but instead drive the model using atmospheric forcings containing projected future climate change signals derived from downscaled Global Climate Model outputs. The SURFEX future simulations will allow investigations of the plausible evolution of Philadelphia's urban climate though the remainder of the 21st century and its future heat stress conditions.

CRediT authorship contribution statement

Benjamin Le Roy: Writing – original draft. **Keith W. Dixon:** Writing – review & editing. **Dennis Adams-Smith:** Writing – review & editing.

Declaration of competing interest

This work was supported by the National Oceanic and Atmospheric Administration's (NOAA) Climate Program Office, NOAA's Geophysical Fluid Dynamics Laboratory, and the Cooperative Institute for Modeling the Earth System - a collaboration between Princeton University and NOAA. The authors declare no conflicts of interest.

Data availability

The SURFEX land surface model used can be downloaded at https://www.umr-cnrm.fr/surfex/. Parts of the data associated with

this study, can be made available on reasonable request.

References

Albergel, C., Dutra, E., Munier, S., Calvet, J.C., Munoz-Sabater, J., de Rosnay, P., Balsamo, G., 2018. ERA-5 and ERA-interim driven ISBA land surface model simulations: which one performs better? Hydrol. Earth Syst. Sci. 22 (6), 3515–3532.

Anderson, G.B., Bell, M.L., Peng, R.D., 2013. Methods to calculate the heat index as an exposure metric in environmental Health Research. Environ. Health Perspect. 121, 1111–1119. https://doi.org/10.1289/ehp.1206273.

Anderson, B., Peng, R., Ferreri, J., 2016. Package 'weathermetrics': functions to convert between weather metrics. Version 1 (2), 2.

Argüeso, D., Evans, J.P., Pitman, A.J., Di Luca, A., 2015. Effects of city expansion on heat stress under climate change conditions. PLoS One 10 (2), e0117066.

Arnfield, A.J., 2003. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. Int. J. Climatol. 23 (1), 1–26.

Baniassadi, A., Sailor, D.J., Ban-Weiss, G.A., 2019. Potential energy and climate benefits of super-cool materials as a rooftop strategy. Urban Clim. 29, 100495.

Bernard, É., De Munck, C., Lemonsu, A., 2022. Detailed mapping and modeling of urban vegetation: what are the benefits for microclimatic simulations with town energy balance (TEB) at neighborhood scale? J. Appl. Meteorol. Climatol. 61 (9), 1159–1178.

Brugha, R., Grigg, J., 2014. Urban air pollution and respiratory infections. Paediatr. Respir. Rev. 15 (2), 194–199.

Bruse, M., 2004. ENVI-met 3.0: updated model overview. University of Bochum (Retrieved from: www. envi-met. com).

Bruse, M., Fleer, H., 1998. Simulating surface–plant–air interactions inside urban environments with a three dimensional numerical model. Environ. Model Softw. 13 (3–4) (384), 373.

Bueno, B., Pigeon, G., Norford, L.K., Zibouche, K., Marchadier, C., 2012. Development and evaluation of a building energy model integrated in the TEB scheme. Geosci. Model Dev. 5 (2), 433–448.

Cannon, A.J., Sobie, S.R., Murdock, T.Q., 2015. Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? J. Clim. 28 (17), 6938–6959.

Chakraborty, T.C., Newman, A.J., Qian, Y., Hsu, A., Sheriff, G., 2023. Residential segregation and outdoor urban moist heat stress disparities in the United States. One Earth 6 (6), 738–750.

Champeaux, J.L., Masson, V., Chauvin, F., 2005. ECOCLIMAP: a global database of land surface parameters at 1 km resolution. Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling 12 (1), 29–32.

Chen, F., Manning, K.W., LeMone, M.A., Trier, S.B., Alfieri, J.G., Roberts, R., Blanken, P.D., 2007. Description and evaluation of the characteristics of the NCAR high-resolution land data assimilation system. J. Appl. Meteorol. Climatol. 46 (6), 694–713.

Chu, E.K., Fry, M.M., Chakraborty, J., Cheong, S.-M., Clavin, C., Coffman, M., Hondula, D.M., Hsu, D., Jennings, V.L., Keenan, J.M., Kosmal, A., Muñoz-Erickson, T.A., Jelks, N.T.O., 2023. Ch. 12. Built environment, urban systems, and cities. In: fifth National Climate Assessment. Crimmins, a.R., C.W. Avery, D.R. Easterling, K.E. Kunkel, B.C. Stewart, and T.K. Maycock, Eds. U.S. global change research program, Washington, DC, USA. https://doi.org/10.7930/NCA5.2023.CH12.

City of Philadelphia, 2015. Building Footprints update. https://opendataphilly.org/datasets/building-footprints/.

Coulston, J.W., Moisen, G.G., Wilson, B.T., Finco, M.V., Cohen, W.B., Brewer, C.K., 2012. Modeling percent tree canopy cover—A pilot study: Photogrammetric Engineering and Remote Sensing, v. 78, no. 7, p. 715–727, at. https://doi.org/10.14358/PERS.78.7.715.

D'Amato, G., Cecchi, L., D'amato, M., Liccardi, G., 2010. Urban air pollution and climate change as environmental risk factors of respiratory allergy: an update. J Investig Allergol Clin Immunol 20 (2), 95–102.

De Munck, C.S., Lemonsu, A., Bouzouidja, R., Masson, V., Claverie, R., 2013. The GREENROOF module (v7. 3) for modelling green roof hydrological and energetic performances within TEB. Geosci. Model Dev. 6 (6), 1941–1960.

Demuzere, M., Hankey, S., Mills, G., Zhang, W., Lu, T., Bechtel, B., 2020. Combining expert and crowd-sourced training data to map urban form and functions for the continental US. Scientific data 7 (1), 264.

Dewitz, J., U.S. Geological Survey, 2021. National Land Cover Database (NLCD) 2019 products (ver. 2.0, June 2021): U.S. Geological Survey data release. https://doi.org/10.5066/P9KZCM54.

Doblas-Reyes, F.J., Sörensson, A.A., Almazroui, M., Dosio, A., Gutowski, W.J., Haarsma, R., Hamdi, R., Hewitson, B., Kwon, W.-T., Lamptey, B.L., Maraun, D., Stephenson, T.S., Takayabu, I., Terray, L., Turner, A., Zuo, Z., 2021. Linking global to regional climate change. 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 [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge university press, Cambridge, United Kingdom and New York, NY, USA, pp. 1363–1512. https://doi.org/10.1017/9781009157896.012.

Duchêne, F., Hamdi, R., Van Schaeybroeck, B., Caluwaerts, S., De Troch, R., De Cruz, L., Termonia, P., 2022. Downscaling ensemble climate projections to urban scale: Brussels's future climate at 1.5° C, 2° C, and 3° C global warming. Urban Clim. 46, 101319.

Falcone, J.A., 2016. U.S. national categorical mapping of building heights by block group from shuttle radar topography Mission data: U.S. Geological Survey data release. https://doi.org/10.5066/F7W09416.

Fan, H., Sailor, D.J., 2005. Modeling the impacts of anthropogenic heating on the urban climate of Philadelphia: a comparison of implementations in two PBL schemes. Atmos. Environ. 39 (1), 73–84.

Geary, R.C., 1936. The distribution of" Student's" ratio for non-normal samples. Suppl. J. R. Stat. Soc. 3 (2), 178–184.

Grell, G.A., Dudhia, J., Stauffer, D.R., 1994. A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5).

Hamdi, R., Masson, V., 2008. Inclusion of a drag approach in the town energy balance (TEB) scheme: offline 1D evaluation in a street canyon. J. Appl. Meteorol. Climatol. 47 (10), 2627–2644.

Hamdi, R., Giot, O., De Troch, R., Deckmyn, A., Termonia, P., 2015. Future climate of Brussels and Paris for the 2050s under the A1B scenario. Urban Clim. 12, 160–182.

Hashemi, F., Poerschke, U., Iulo, L.D., Chi, G., 2023. Urban microclimate, outdoor thermal comfort, and socio-economic mapping: a Case study of Philadelphia. PA. Buildings 13 (4), 1040.

Hayden, M.H., Schramm, P.J., Beard, C.B., Bell, J.E., Bernstein, A.S., Bieniek-Tobasco, A., Cooley, N., Diuk-Wasser, M., Dorsey, Michael K., Ebi, K.L., Ernst, K.C., Gorris, M.E., Howe, P.D., Khan, A.S., Lefthand-Begay, C., Maldonado, J., Saha, S., Shafiei, F., Vaidyanathan, A., Wilhelmi, O.V., 2023. Ch. 15. Human health. In: Fifth National Climate Assessment. Crimmins, A.R., C.W. Avery, D.R. Easterling, K.E. Kunkel, B.C. Stewart, and T.K. Maycock, Eds. U.S. Global Change Research Program, Washington, DC, USA. https://doi.org/10.7930/NCA5.2023.CH15.

Heaviside, C., Macintyre, H., Vardoulakis, S., 2017. The urban heat island: implications for health in a changing environment. Current environmental health reports 4, 296–305.

Heris, M.P., Foks, N.L., Bagstad, K.J., Troy, A., Ancona, Z.H., 2020. A rasterized building footprint dataset for the United States. Scientific data 7 (1), 207.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, J.N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146 (730), 1999–2049.

Kusaka, H., Kondo, H., Kikegawa, Y., Kimura, F., 2001. A simple single-layer urban canopy model for atmospheric models: comparison with multi-layer and slab models. Bound.-Layer Meteorol. 101, 329–358.

Landsberg, H.E., 1970, October. Meteorological observations in urban areas. In: Meteorological Observations and Instrumentation: Proceedings of the American Meteorological Society Symposium on Meteorological Observations and Instrumentation, Washington, DC, 10–14 February 1969. American Meteorological Society, Boston, MA, pp. 91–99.

Langendijk, G.S., Rechid, D., Jacob, D., 2019. Urban areas and urban–rural contrasts under climate change: what does the EURO-CORDEX ensemble tell us?—investigating near surface humidity in Berlin and its surroundings. Atmosphere 10 (12), 730.
Lanzante, J.R., Adams-Smith, D., Dixon, K.W., Nath, M., Whitlock, C.E., 2020. Evaluation of some distributional downscaling methods as applied to daily maximum temperature with emphasis on extremes. Int. J. Climatol. 40 (3), 1571–1585.
Le Roy, B., Lemonsu, A., Schoetter, R., 2021. A statistical–dynamical downscaling methodology for the urban heat island applied to the EURO-CORDEX ensemble. Clim. Dyn. 56, 2487–2508.
Lemonsu, A., Masson, V., Shashua-Bar, L., Erell, E., Pearlmutter, D., 2012. Inclusion of vegetation in the town energy balance model for modelling urban green areas. Geosci. Model Dev. 5 (6), 1377–1393.
Le Roy, B., Lemonsu, A., Schoetter, R., Machado, T., 2024. Study of the future evolution of the urban climate of Paris by statistical-dynamical downscaling of the EURO-CORDEX ensemble. J. Appl. Meteorol. Climatol.
Lemonsu, A., Kounkou-Arnaud, R., Desplat, J., Salagnac, J.L., Masson, V., 2013. Evolution of the Parisian urban climate under a global changing climate. Clim. Chang. 116, 679–692.
Lu, Y.-C., Romps, D.M., 2022. Extending the heat index. J. Appl. Meteorol. Climatol. 61, 1367–1383. https://doi.org/10.1175/JAMC-D-22-0021.1.
Masson, V., 2000. A physically-based scheme for the urban energy budget in atmospheric models. Boundary-layer meteorology 94, 357–397.
Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Voldoire, A., 2013. The SURFEXv7. 2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. Geosci. Model Dev. 6 (4), 929–960.
Masson, V., Marchadier, C., Adolphe, L., Aguejdad, R., Avner, P., Bonhomme, M., Zibouche, K., 2014. Adapting cities to climate change: a systemic modelling approach. Urban Clim. 10, 407–429.
Masson, V., Lemonsu, A., Hidalgo, J., Voogt, J., 2020. Urban climates and climate change. Annu. Rev. Environ. Resour. 45, 411–444.
Mironov, D.V., 2008. Parameterization of Lakes in Numerical Weather Prediction: Description of a Lake Model (pp. 41-pp). Offenbach, Germany: DWD.
Nogueira, M., Albergel, C., Boussetta, S., Johannsen, F., Trigo, I.F., Ermida, S.L., Dutra, E., 2020a. Role of vegetation in representing land surface temperature in the CHTESSEL (CY45R1) and SURFEX-ISBA (v8. 1) land surface models: a case study over Iberia. Geosci. Model Dev. 13 (9), 3975–3993.
Nogueira, M., Lima, D.C., Soares, P.M., 2020b. An integrated approach to project the future urban climate response: changes to Lisbon's urban heat island and temperature extremes. Urban Clim. 34, 100683.
Nogueira, M., Hurduc, A., Ermida, S., Lima, D.C., Soares, P.M., Johannsen, F., Dutra, E., 2022. Assessment of the Paris urban heat island in ERA5 and offline SURFEX-TEB (v8. 1) simulations using the METEOSAT land surface temperature product. Geosci. Model Dev. 15 (14), 5949–5965.
Noilhan, J., Planton, S., 1989. A simple parameterization of land surface processes for meteorological models. Mon. Weather Rev. 117 (3), 536–549.
Nunez, M., Oke, T.R., 1977. The energy balance of an urban canyon. J. Appl. Meteorol. Climatol. 16 (1), 11–19.
Oke, T.R., 2004. Urban Observations. World Meteorological Organization, IOM Report, p. 81.
Oke, T.R., Mills, G., Christen, A., Voogt, J.A., 2017. Urban climates. Cambridge University Press.
Oleson, K.W., Monaghan, A., Wilhelmi, O., Barlage, M., Brunsell, N., Feddema, J., Steinhoff, D.F., 2015. Interactions between urbanization, heat stress, and climate change. Clim. Chang. 129, 525–541.
Otte, T.L., Lacser, A., Dupont, S., Ching, J.K., 2004. Implementation of an urban canopy parameterization in a mesoscale meteorological model. J. Appl. Meteorol. 43 (11), 1648–1665.
Pearson, K., 1905. Skew variation, a rejoinder. Biometrika 4 (1–2), 169–212.
Pekel, J.F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540 (7633), 418–422.
Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Hofton, M., 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sens. Environ. 253, 112165.
R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
Ramamurthy, P., Bou-Zeid, E., 2017. Heatwaves and urban heat islands: a comparative analysis of multiple cities. J. Geophys. Res. Atmos. 122 (1), 168–178.
Ranasinghe, R., Ruane, A.C., Vautard, R., Arnell, N., Coppola, E., Cruz, F.A., Zaaboul, R., 2021. Climate change information for regional impact and for risk assessment.
Redon, E., Lemonsu, A., Masson, V., 2020. An urban trees parameterization for modeling microclimatic variables and thermal comfort conditions at street level with the town energy balance model (TEB-SURFEX v8. 0). Geosci. Model Dev. 13 (2), 385–399.
Rizwan, A.M., Dennis, L.Y., Chunho, L.I.U., 2008. A review on the generation, determination and mitigation of urban Heat Island. J. Environ. Sci. 20 (1), 120–128.
Rothfusz, L.P., Headquarters, N.S.R., 1990. The heat index equation (or, more than you ever wanted to know about heat index). Fort Worth, Texas: National Oceanic and Atmospheric Administration, National Weather Service, Office of Meteorology 9023, 640.
Salgado, R., Le Moigne, P., 2010. Coupling of the FLake Model to the Surfex Externalized Surface Model.
Schmidt, G.A., Ruedy, R., Hansen, J.E., Aleinov, I., Bell, N., Bauer, M., Yao, M.S., 2006. Present-day atmospheric simulations using GISS ModelE: comparison to in situ, satellite, and reanalysis data. J. Clim. 19 (2), 153–192.
Schmidt, G.A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G.L., Aleinov, I., Zhang, J., 2014. Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. Journal of Advances in Modeling Earth Systems 6 (1), 141–184.
Shandas, V., Voelkel, J., Williams, J., Hoffman, J., 2019. Integrating satellite and ground measurements for predicting locations of extreme urban heat. Climate 7, 5. https://doi.org/10.3390/cli7010005.
Skamarock, W.C., Klemp, J.B., 2008. A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. J. Comput. Phys. 227 (7), 3465–3485.
Soltani, A., Sharifi, E., 2017. Daily variation of urban heat island effect and its correlations to urban greenery: a case study of Adelaide. Frontiers of Architectural Research 6 (4), 529–538.
Steadman, R.G., 1979. The assessment of sultriness. Part I: a temperature-humidity index based on human physiology and clothing science. J. Appl. Meteorol. Climatol. 18 (7), 861–873.
Strategies, C.A.P.A., 2022. Heat Watch Philadelphia Report. https://osf.io/3wcmr.
Trail, M., Tsimpidi, A.P., Liu, P., Tsigaridis, K., Hu, Y., Nenes, A., Russell, A.G., 2013. Downscaling a global climate model to simulate climate change over the US and the implication on regional and urban air quality. Geosci. Model Dev. 6 (5), 1429–1445.
Wang, Z.H., Bou-Zeid, E., Smith, J.A., 2013. A coupled energy transport and hydrological model for urban canopies evaluated using a wireless sensor network. Q. J. R. Meteorol. Soc. 139 (675), 1643–1657.
Xian, G., Homer, C., Yang, L., 2011. Development of the USGS National Land-Cover Database over two decades. Advances in Environmental Remote Sensing 525.
Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Grannemann, B., Rigge, M., Xian, G., 2018. A new generation of the United States National Land Cover Database: requirements, research priorities, design, and implementation Strategies. ISPRS J. Photogramm. Remote Sens. 146, 108–123.
Yoo, C., Im, J., Park, S., Quackenbush, L.J., 2018. Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data. ISPRS J. Photogramm. Remote Sens. 137, 149–162. https://doi.org/10.1016/j.isprsjprs.2018.01.018.