High-resolution global climate simulations: representation of cities

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See .

To link to this article DOI: http://dx.doi.org/10.1002/joc.7018

RESEARCH ARTICLE

High-resolution global climate simulations: Representation of cities

Denise Hertwig[1] | **Matthew Ng[1]** | **Sue Grimmond[1]** | **Pier Luigi Vidale[1,2]** |
Patrick C. McGuire[1,2,3]

[1]Department of Meteorology, University of Reading, Reading, UK

[2]National Centre for Atmospheric Science, Reading, UK

[3]Department of Geography and Environmental Science, University of Reading, Reading, UK

Correspondence to: Denise Hertwig, Department of Meteorology, University of Reading, P.O. Box 243, Reading, Berkshire, RG6 6ET, UK.
E-mail: d.hertwig@reading.ac.uk

Abstract

Ensemble runs of high-resolution ($\sim$10 km; N1280) global climate simulations (2005–2010) with the Met Office HadGEM3 model are analysed over large urban areas in the south-east UK (London) and south-east China (Shanghai, Hangzhou, Nanjing region). With a focus on urban areas, we compare meteorological observations to study the response of modelled surface heat fluxes and screen-level temperatures to urbanization. HadGEM3 has a simple urban slab scheme with prescribed, globally fixed bulk parameters. Misrepresenting the magnitude or the extent of urban land cover can result in land-surface model bias. As urban land-cover fractions are severely under-estimated in China, this impacts surface heat-flux partitioning and quintessential features, such as the urban heat island. Combined with the neglect of anthropogenic heat emissions, this can result in misrepresentation of heat-wave intensities (or cold spells) in cities. The model performance in urban areas could be improved if bulk parameters are modelled instead of prescribed, but this necessitates the availability of local morphology data on a global level. Improving land-cover information and providing more flexible ways to account for differences between cities (e.g., anthropogenic emission; morphology) is essential for realistic future projections of city climates, especially if model output is intended for urban climate services.

KEYWORDS

cities, global climate simulation, screen-level temperature, urbanization, urban climate, urban land-surface model

1 | INTRODUCTION

Cities are key to global climate change mitigation strategies (e.g., Mi *et al.*, 2019 and references therein), and, at the same time, have to respond to the local impacts of climate change through adaptation measures (e.g., Landauer *et al.*, 2019). Rapidly growing urban populations (UN, 2019) increase the need to conduct projections of

future city climates to inform and develop integrated urban climate services, for example, to provide guidance for sustainable urban adaptation and planning (e.g., Cortekar *et al.*, 2016; Baklanov *et al.*, 2018; Grimmond *et al.*, 2020).

The volume and density of buildings and the thermal and radiative properties of urban materials affect the surface-energy balance in cities, for example, through augmentation of heat storage, enhanced radiative

trapping and increased aerodynamic roughness. A well-studied manifestation of this is the urban heat island (Oke, 1973), that is, warmer cities compared to their rural surroundings. The prevalence of impervious surfaces also affects the water balance in cities, with increased surface runoff and strongly reduced potential for infiltration and evaporation (Grimmond, 2007).

Such localized impacts of urbanization can be represented in atmospheric models through the use of specialized urban schemes in land-surface models. Urban land-surface models (ULSM) parameterise the urban surface energy balance (e.g., Oke *et al.*, 2017),

$$Q_N + Q_F = Q_H + Q_E + \Delta Q_S, \qquad (1)$$

where the net all-wave radiation, $Q_N = (K^\downarrow - K^\uparrow) + (L^\downarrow - L^\uparrow)$, is driven by the net-radiative forcing from incoming ($^\downarrow$) and outgoing ($^\uparrow$) shortwave (K) and longwave (L) radiation. Additional energy input from anthropogenic heat emissions (Q_F) depends on factors such as population density, traffic volume and seasonal heating/cooling demands, and hence can vary strongly in space and time (e.g., Sailor, 2011; Iamarino *et al.*, 2012; Gabey *et al.*, 2019). In some situations, Q_F can be the dominant energy source in Equation (1), for example, in mid-latitude cities in winter when Q_N is low and space heating widely used (e.g., Hamilton *et al.*, 2009). A portion of the energy received is stored (ΔQ_S) in the urban volume (e.g., walls, streets) and the ground. The remaining energy, partitioned into turbulent sensible (Q_H) and latent (Q_E) heat fluxes, is essential to drive boundary-layer dynamics. When the ULSM is coupled to a large-scale atmospheric model, the net horizontal advection of energy into the system (ΔQ_A) is accounted for in Equation (1) (right-hand side).

With increasing resolution of both regional and global climate models, cities and conurbations start to become resolvable on the model grid (if using adequate land-cover information), so that local and regional effects of urbanization can, in principle, be represented in climate simulations (e.g., Oleson *et al.*, 2008a, 2008b). However, global simulations still rarely model urban effects explicitly (*cf.* discussions in Daniel *et al.*, 2019). This requires not only adequate urban land-cover information, but for more complex urban models also the availability of bulk information on building morphology (e.g., mean building height) and anthropogenic heat emissions (Grimmond *et al.*, 2009), which are not readily available at the global scale.

In this study, we address the following questions:

- How well do simulations with a simple urban (bulk) scheme of recent climate (2005–2010) agree with surface observations of screen-level temperatures?

- What are the main factors affecting the representation of urban signals in these simulations?
- What data/modelling capabilities are needed to improve the representation of urban signatures in climate projections and to provide integrated urban climate services for future planning scenarios?

Data from high-resolution (∼10 km; N1280) global climate simulations with the Met Office HadGEM3 model (Section 2.1) are analysed over highly urbanized and populous metropolitan regions (Section 2.2): (a) the south-east UK (London) and (b) south-east China (Yangtze River Delta region including Shanghai, Hangzhou and Nanjing).

2 | CLIMATE SIMULATIONS

2.1 | Setup

The frontier climate simulations of the H2020 PRIMAVERA project (Vidale *et al.*, in prep.), with a global horizontal resolution of ∼0.1° (∼10 km; N1280), use the Met Office HadGEM3 climate modelling system (Hewitt *et al.*, 2011). The atmosphere-only simulations for a recent 6-year period (2005–2010) use the Global Atmosphere (GA7.1) and Global Land (GL7.0) science configurations for the Met Office Unified Model (UM; Walters *et al.*, 2019, Wiltshire *et al.*, 2020) and the Joint UK Land Environment Simulator (JULES; Best *et al.*, 2011, Clark *et al.* 2011). Sea surface temperatures and sea ice concentrations are used as boundary conditions. These are obtained from the HadISST2.1.1 dataset (Titchner and Rayner, 2014) and are modified to provide daily increments suitable for high-resolution simulations.

Given the large computational costs of the simulation, the model is initialised from a converged state at coarser horizontal resolution (25 km, N512; Roberts *et al.*, 2019). To help eliminate sensitivity to initial conditions, an ensemble of high-resolution simulations is constructed. We analyse three ensemble members, labelled HAD_{1-3}, available from the high-resolution HadGEM3–PRIMAVERA N1280 suite. The ensemble members are spawned by randomly perturbing the initial conditions of the surface temperature field and by applying to each member a further stochastic perturbation to the full set of initial and restart conditions as they are read in. Stochastic perturbations are needed as the model uses stochastic physics. These are additionally injected at regular intervals during the simulation, with different time scales depending on the spatial scale (Sanchez *et al.*, 2016).

The JULES land-surface model employs a tile approach for the sub-grid-scale heterogeneity of the land surface in a grid-box (Best *et al.*, 2011). In the JULES–GL7.0 science configuration, the built part (i.e., buildings, roads) of urban areas is represented by a single tile (Best–1T model; Best, 2005) with only one globally constant set of surface parameter values for heat capacity, albedo, emissivity and surface roughness. Thus, they do not reflect morphometric differences between cities and across a city (e.g., Grimmond and Oke, 1999; Kent *et al.*, 2019). The values used are (Wiltshire *et al.*, 2020): aerodynamic roughness length of $z_o = 1$ m, bulk urban albedo of $\alpha = 0.18$, bulk urban emissivity of $\varepsilon = 0.97$ and surface heat capacity of $C = 0.28$ MJ K^{-1} m^{-2}. The ratio of roughness length for heat (z_h) and momentum (z_o) remains constant at 10^{-7} (Best *et al.*, 2006). As buildings are assumed to be below the surface, a zero-plane displacement is assumed to be unneeded (i.e., 0 m).

In these simulations, irrigation or other anthropogenic moisture sources are not modelled for urban or vegetation tiles. Following precipitation, the built surface sheds most of the water immediately (i.e., urban run-off rates are large; Hertwig *et al.*, 2020) as the water-holding capacity of the impervious tiles is limited (Best and Grimmond, 2016a). Hence, in most situations (i.e., in the absence of rainfall), the turbulent latent heat flux (Q_E) is limited to the vegetated tiles, which did not receive any urban runoff. Anthropogenic heat emissions (Q_F) in the HadGEM3–PRIMAVERA runs are assumed to be 0 W m^{-2} everywhere. JULES allows Q_F to be prescribed monthly for urban tiles, but these values are spatially unvarying as, for example, used in UK operational numerical weather prediction with the UM (Lean *et al.*, 2011). While global Q_F models exist (e.g., Allen *et al.*, 2011; Lindberg *et al.*, 2013), JULES currently only allows the anthropogenic forcing to be prescribed, which requires access to local energy-use statistics. These are difficult to obtain with appropriate timeliness at the global scale. Hence, including Q_F in global HadGEM3–JULES climate model runs is not currently viable.

Vegetation in HadGEM3-JULES is represented using five plant functional types (PFT): broad-leaf tree, needle-leaf tree, shrubs, C3 and C4 grass. Urban vegetation (if resolved) is modelled with one of these vegetation tiles, but without representing interactions with impervious surfaces. Leaf-area index and canopy height for the vegetation tiles are prescribed as monthly values that vary at the grid-scale (Wiltshire *et al.*, 2020). Neither crop dynamics nor its irrigation are modelled.

Evaluations of HadGEM3 (coupled to an ocean model, Roberts *et al.*, 2019; and/or atmosphere-only, Vannière *et al.*, 2019) have shown that increasing horizontal resolution from 130 (N96) to 25 km (N512)

reduces global biases of near-surface temperatures and improves rainfall patterns and amplitudes. Detailed evaluations of the HadGEM3–PRIMAVERA N1280 runs are ongoing (Vidale *et al.*, in prep.; Volonté, pers. comm.) as the model continues to be developed. Biases identified at 25-km model resolution are not expected to fundamentally shift with the change to 10 km, consistent with Vellinga *et al.* (2016) for the switch from 25 to 12 km.

2.2 | Analysis domains

Model output and surface observations are compared for domains centred on the metropolitan areas of London (Figure 1a,b) and Shanghai (Figure 1c,d). As London is central to development and evaluation of the urban models in JULES (Best, 2005; Porson *et al.*, 2010; Bohnenstengel *et al.*, 2011), the urban parameters used in HadGEM3–PRIMAVERA are expected to be most representative of this urban area.

The $\sim$1 km resolution International Geosphere–Biosphere Project (IGBP; Loveland *et al.*, 2000) data for 1995/1996 are used to create the GA7.1/GL7.0 model land cover (Walters *et al.*, 2019). Comparison of the urban impervious or built land-cover fraction (f_{Urban}) used in HadGEM3–PRIMAVERA (Figure 1b,d) to more recent (2011) data from the Global Urban Footprint (GUF) project ($\sim$12 m resolution globally; Esch *et al.*, 2011, 2017, 2018) using TerraSAR–X/TanDEM–X data (Figure 1a,c) allows differences to be seen.

As London is the largest metropolitan area in Europe, it covers several HadGEM3 grid cells (Figure 1b). In central London, the urban land-cover fraction is at or close to 100% (i.e., vegetation and the River Thames are missing; Table 1a). Shanghai, China's most populous city, and surrounding regions of the Yangtze Estuary have undergone rapid urbanization over the last decades (Yin *et al.*, 2011; Cui and Shi, 2012; Tan *et al.*, 2015). However, this is not evident in the HadGEM3 land cover. The vast cities of Shanghai and Hangzhou (Figure 1c), for example, are only represented by few urbanized grid-boxes (Figure 1d). The urban fraction is too low compared to GUF data (Table 1b), with the maximum f_{Urban} across the domain being 60.2% (inner-city Shanghai). In both domains, C3 grass is the dominant PFT with 90% of grid-boxes having tile-fractions larger than 59% (south-east UK) or 46% (south-east China), while less than 10% of the grid-boxes have trees.

Similarly, large contrasts are evident between IGBP f_{Urban} and GUF in other highly urbanized and populous areas across China (Figure S1 for Beijing and Chongqing). Given this clear bias in China, this study focuses only on a small-area comparison centred on Shanghai and cities located in the lower reaches of the Yangtze
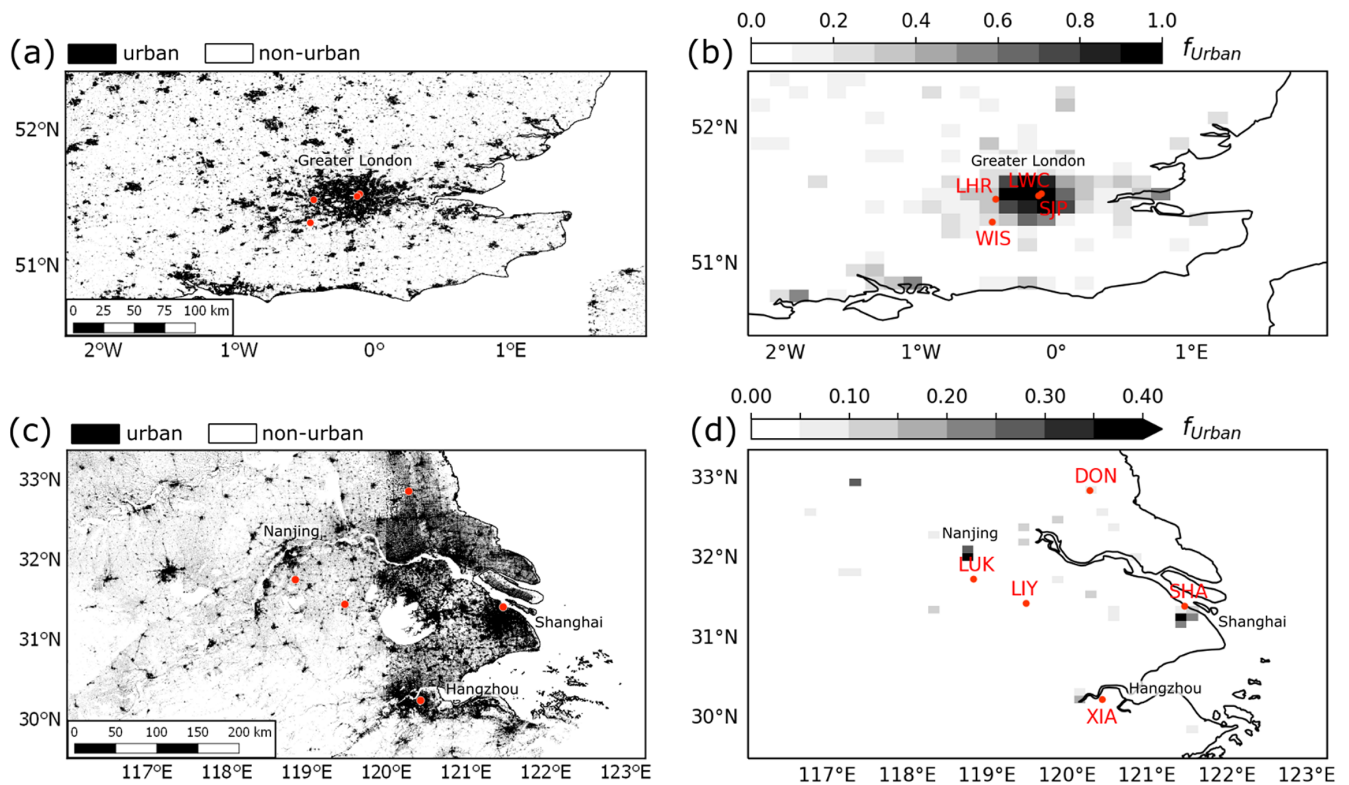
FIGURE 1 Domains analysed with observation sites (dots, labels) and the (a,c) urban and non-urban extent of the Global Urban Footprint (GUF) dataset compared to the (b,d) HadGEM3–PRIMAVERA urban land-cover fractions (f_{Urban}) for (a,b) south-east UK, (c,d) Shanghai region in China. Note the f_{Urban} scale differs between (b) and (d). Table 1 and text provide further details and data sources [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Weather station (Figure 1b,d) data analysed in two domains located in (a) south-east UK (London region) and (b) south-east China (Shanghai region) with surrounding urban land-cover fraction (f_{Urban}) determined from GUF (25 and 2.5 km boxes centred on the sites; *cf.* Figures S2 and S3) and HadGEM3–PRIMAVERA (3-by-3 grid-boxes ≈ 30 km horizontal extent; minimum, maximum and centre values)

| | Station | Latitude | Longitude | f_{Urban} (%) | | | Comment |
				GUF$_{25 km}$	GUF$_{2.5 km}$	HadGEM3 (min/max/centre)	
(a) London region	LWC	51.52°	−0.11°	75.5	90.7	45.2/100/97.7	stopped 02/2010
	SJP	51.50°	−0.13°	75.6	80.6	45.2/100/97.7	
	LHR	51.48°	−0.45°	44.0	48.3	7.0/91.5/33.1	Heathrow airport
	WIS	51.31°	−0.74°	23.1	3.5	7.1/75.4/26.1	
(b) Shanghai region	SHA	31.4°	121.47°	43.9	87.3	0/60.2/6.3	Baoshan district
	XIA	30.23°	120.43°	39.4	35.4	0/0/0	Hangzhou airport
	DON	32.85°	120.28°	20.8	36.1	0/9.7/9.7	
	LIY	31.43°	119.48°	13.2	81.9	0/0/0	
	LUK	31.74°	118.86°	13.2	19.1	0/0/0	Nanjing/Lukou airport

Note: See text for more details. Stations are: DON, Dongtai (WMO: 58251); LIY, Liyang (WMO: 58345); LHR, Heathrow Airport (WMO: 03772); LWC, London Weather Centre (WMO code: 03779); LUK, Lukou (ICAO code: ZSNJ); SHA, Shanghai (WMO: 58362); SJP, St James's Park (WMO: 03770); WIS, Wisley (DCNN code: 5237); XIA, Xiaoshan (WMO: 58457).

River Delta, for which the observational coverage is good (Section 3).

2.3 | Output

Surface turbulent sensible and latent heat fluxes, Q_H and Q_E, and net all-wave radiation (Q_N) are available as daily averages from three ensemble members (HAD$_{1-3}$). These are analysed for the entire 6-year simulation period (2005–2010). Hourly samples of air temperature (T_{air}) at screen-height (1.5 m above local orography) are available from two ensemble members (HAD$_{1,2}$). T_{air} is a model diagnostic obtained from interpolating air temperatures between the first model level and the surface using Monin–Obukhov similarity theory (Essery *et al.*, 2001; Bohnenstengel *et al.*, 2014). As ensemble member HAD$_2$ has no output for 2010, the seasonal and diurnal comparisons are only for 5 years (2005–2009).

Given the model years have only 360 days, the comparisons with observations (Section 3) cannot be for individual short periods (e.g., hourly) but need to be based on aggregated data (e.g., monthly, seasonal; Section 4.2) and use frequency distributions and occurrence of particular conditions (e.g., hot/cold days; Section 4.3). Comparison of model statistics of T_{air} with the observations (Section 3) are conducted for common data periods. Frequency distributions (Section 4.3) use the same sample frequency for model and observations; that is, the hourly model output is sampled to match the 3-hourly observation frequency available for sites in the Shanghai domain (for the UK sites it is hourly; see Section 3). Data are analysed by month and season: summer (June, July, August; JJA), autumn (SON), winter (DJF), and spring (MAM). The model bias is analysed as the difference between model and observations.

3 | OBSERVATIONS

Evaluation of modelled T_{air} uses single-site surface observations from weather stations. For the south-east UK, hourly samples of screen-level air temperature from the Met Office MIDAS surface observations archive (Met Office, 2006) are used at four sites (Figure 1b; Table 1a). In the south-east China domain, ground-based observations from NOAA's National Climatic Data Center (NCDC) archive (Climate Data Online; https://www7.ncdc.noaa.gov/CDO/cdo) at five sites (Figure 1d, Table 1b) are selected based on data availability during the evaluation period. From NCDC, 3-hourly T_{air} samples are available at 0, 3, ..., 21 UTC. Both MIDAS and NCDC T_{air} data have a resolution of 0.1°C.

Urban land-cover characteristics of the surface stations are summarized (Table 1) using both GUF (25^2 and 2.5^2 km^2 footprints around the sites) and HadGEM3 land cover (Section 2.2). The latter is given for a 3-by-3 grid-box area ($\sim30^2$ km^2) centred on the evaluation sites. For some stations, the GUF$_{2.5\ km}$ and GUF$_{25\ km}$ differences are large (Figures S2, S3). For example, the immediate surroundings of WIS (UK) are primarily rural ($f_{Urban} = 3.5\%$) compared to the wider regional average (23.1%). The latter is close to the HadGEM3 value (26.1%).

The opposite occurs for LIY (China), where GUF$_{25\ km}$ indicates settlements in rural surroundings, while the site's immediate neighbourhood is highly urbanized (81.9%). For SHA (Baoshan/Shanghai), the close proximity to the Yangtze River causes GUF$_{25\ km}$ to be low, whereas GUF$_{2.5\ km}$ ($f_{Urban} = 87.3\%$) indicates extensive urban surroundings to the station, and of the inner-city area south of the site. However, f_{Urban} in the HadGEM3 land cover is zero in the 3-by-3 grid-box area around the LIY, XIA and LUK sites (China), in strong contrast to GUF (Table 1b).

4 | RESULTS AND DISCUSSION

4.1 | Surface forcing

Before evaluating screen-level temperatures (Section 4.2, 4.3), modelled turbulent heat fluxes at the surface from the three model ensemble members (HAD$_{1-3}$) are analysed, as these directly impact the boundary-layer temperature profiles. The nature of the energy partitioning into sensible and latent turbulent heat flux (Equation 1) is often strongly altered in cities compared to rural areas in response to the prevalence of impervious surfaces and reduced amounts of vegetation (e.g., Grimmond and Oke, 2002; Goldbach and Kuttler, 2013; Ward *et al.*, 2016). The relative dominance of Q_H or Q_E for a site can be assessed by the evaporative fraction

$$EF = Q_E/(Q_E + Q_H) = 1/(1+\beta) \tag{2}$$

where $\beta = Q_H/Q_E$ is the Bowen ratio, another widely used method to characterize surface energy-flux partitioning.

The ensemble median summer (JJA; 2005–2010) surface heat fluxes are derived from the daily-averaged model fluxes of all three ensemble members (HAD$_{1-3}$). Spatial distributions of Q_H, Q_E and EF are analysed in the south-east UK domain (Figure 2a-c), as the HadGEM3 land-cover data here are generally reasonable for f_{Urban} (Table 1a). Patterns of enhanced (reduced) JJA Q_H (Q_E) are clearly linked to urban fractions for individual
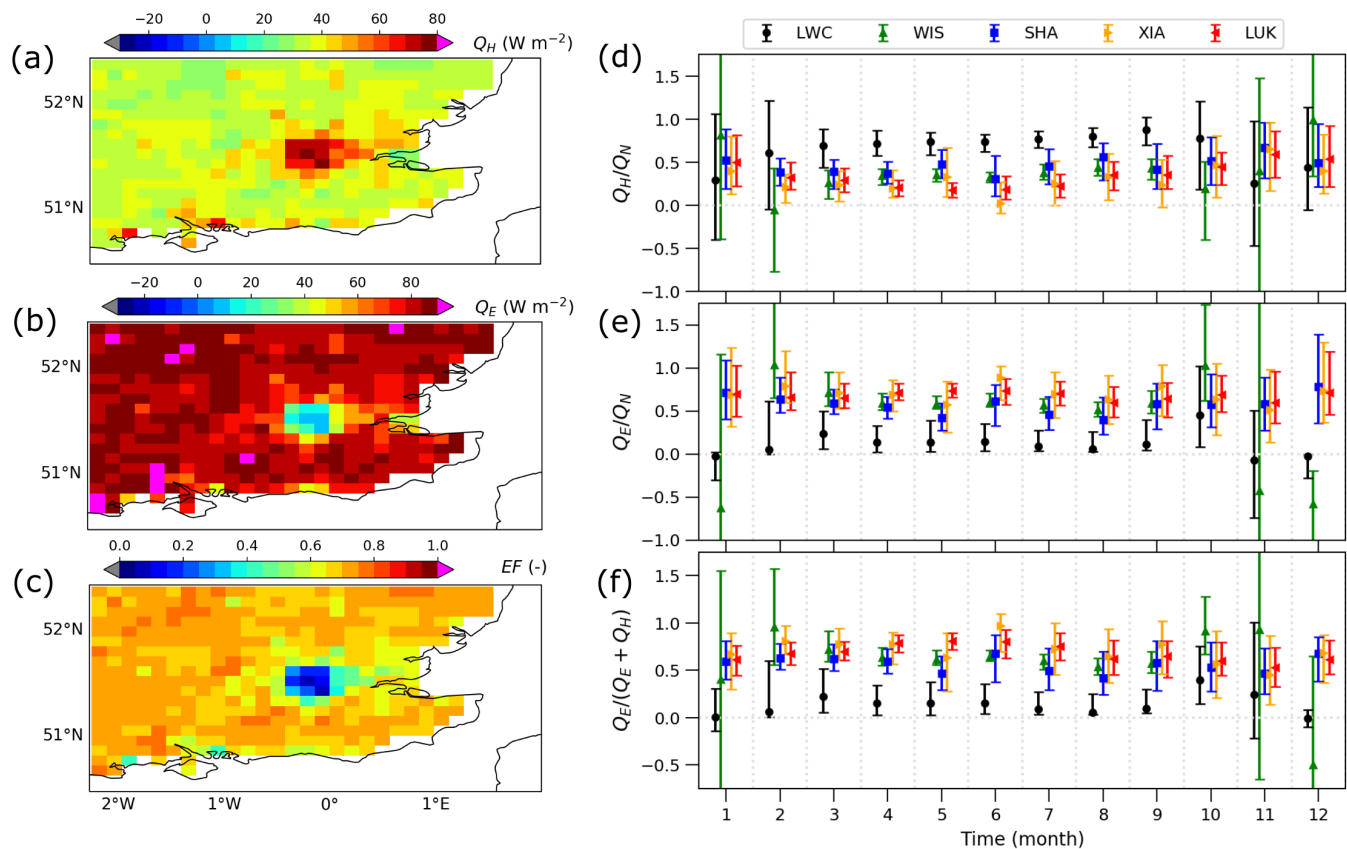
FIGURE 2 (a–c) Modelled summer (JJA; 2005–2010) ensemble median (HAD$_{1-3}$) turbulent surface heat fluxes over land for the south-east UK of (a) Q_H, (b) Q_E and (c) evaporative fraction (*EF*; Equation 2). Monthly ensemble medians (HAD$_{1-3}$, 2005–2010) with inter-quartile range (whiskers) for (d) Q_H/Q_N, (e) Q_E/Q_N and (c) $EF = Q_E/(Q_E + Q_H)$ at LWC, WIS (south-east UK domain; Table 1a) as well as SHA, XIA, and LUK (south-east China domain; Table 1b). Note these analyses use daily mean Q_H, Q_E and Q_N [Colour figure can be viewed at wileyonlinelibrary.com]

HadGEM3 surface grid-boxes (Figure 1b). The Greater London area stands out distinctly from its rural surroundings. An observed heat-flux climatology for central London (Kotthaus and Grimmond, 2014a) informs expectations of model behaviour. With only daily-mean model fluxes available, diurnal variability is unknown and the magnitudes are small (*cf.* hourly-mean observations). In central London, the observed JJA 25th and 75th percentiles of Q_H at the time of day when the median Q_N is largest can be between ∼150–250 W m^{-2} and at night between ∼50–100 W m^{-2} when Q_N is lowest and negative (see Figure 6 in Kotthaus and Grimmond, 2014a). Hence, the JJA model median of $Q_H < 80$ W m^{-2} in central London is lower than expected.

In grid-boxes with $f_{Urban} > 90\%$ in central London (Figure 1b), Q_E predominantly occurs immediately after rainfall, but JULES eliminates subsequent urban tile evaporation through large runoff rates, which results in JJA median daily-mean Q_E of ≤10 W m^{-2} (Figure 2b). Earlier studies demonstrate that even small amounts of vegetation in cities can have a strong impact on the local surface energy balance (e.g., Loridan and

Grimmond, 2012a, 2012b; Best and Grimmond, 2016b). The steep spatial gradients of Q_E (Figure 2b) as f_{Urban} decreases away from central London demonstrate this, as well. Like Q_F, anthropogenic moisture sources are not modelled in JULES for urban tiles. However, these can be important in some cities. Recent studies in Beijing (Dou *et al.*, 2019) and Shanghai (Ao *et al.*, 2018) demonstrate that irrigation of vegetation, street cleaning and/or wetting (e.g., to reduce dust, to cool) have a noticeable, non-negligible effect on observed Q_E.

The median JJA evaporative fractions (Equation 2; Figure 2c) for the model grid-boxes covering central London are clearly dominated by Q_H with values between $0.075 < EF < 0.3$. Hence, the corresponding Bowen ratio (β) is between 12 and 2.3, which is high, but not unreasonable for central London. Kotthaus and Grimmond (2014a) report monthly median hourly β between 5 and 10 in London's central business district. In rural grid-boxes ($f_{Urban} = 0$), modelled EF is larger ($0.65 < EF < 0.75$; i.e. β ≈ 0.54–0.33) and more spatially homogeneous.

To illustrate the sensitivity of JULES surface fluxes to land cover and urban parameter choices, JULES is run offline mimicking the HadGEM3–PRIMAVERA GL7.0 setup at a site in central London (KCL) for which high-quality flux measurements are available for 3 years (2011–2013; Ward *et al.*, 2016, Hertwig *et al.*, 2020). A description of JULES settings, parameters (Table A1) and observations are given in Appendix A. The response of the Best–1T urban model (Section 2.1) is tested in two configurations: (a) a control case using high-resolution land-cover data, observed roughness and radiative parameters and realistic (modelled) Q_F (hereafter CTRL–1T), and (b) default JULES–GL7.0 parameters (Section 2.1) and HadGEM3–PRIMAVERA land cover (HAD–1T).

Seasonal median diurnal cycles of Q_H modelled offline (Figure 3a) show distinct differences in the two test cases, highlighting some probable causes for the underprediction of the (online) HadGEM3 Q_H (despite the over-prediction of f_{Urban}) in central London (Figure 2a). Except for afternoons and early evenings in MAM and JJA, both CTRL–1T and HAD–1T have a consistently negative Q_H bias. Neglecting Q_F in HAD–1T noticeably amplifies this, especially in SON and DJF (Figure 3a), where there are long periods of median Q_H at or below 0 W m^{-2}. The observations instead show $Q_H > 0$ W m^{-2} year-round, day and night (the observed DJF minimum 25th percentile of Q_H is ∼25 W m^{-2}), attributed to both Q_F and nocturnal storage heat flux from the building volume. Underestimation of Q_H is also

caused by the relatively large JULES–GL7.0 urban albedo used in HAD–1T ($\alpha = 0.18$) compared to the observed value of 0.11 used in CTRL–1T (Table A1), leading to a reduction of energy input (Q_N) through an increase of $K^{\uparrow}$ (Figure S4a,c). In both configurations, Q_H has a substantial phase delay (rise and peak times), as a result of the large thermal inertia (through C and z_h) of the urban slab impacting the temporal response of surface temperatures (Hertwig *et al.*, 2020). The phase delay, also present in the diurnal cycle of $L^{\uparrow}$ (Figure S4b), impacts diagnostics like T_{air} (Section 4.2).

In CTRL–1T (Figure 3b), vegetation (13% land-cover fraction within a 500-m radius around the KCL site; Table A1) and the River Thames (21%) influence Q_E, while in HAD–1T there is no vegetation and only 2% water. This results in a median Q_E close to 0 W m^{-2} in HAD–1T (Figure 3b). This echoes the patterns of very low Q_E in central London in the HadGEM3–PRIMAVERA simulations (Figure 2b,e). The river influences the eddy-covariance heat-flux source areas only for some wind directions (Kotthaus and Grimmond, 2014b). As the relative location of land cover is not captured in the JULES tiling, CTRL–1T overestimates the effect of the river on Q_E, which in turn leads to reduced Q_H (Figure 3a).

The impact of over- or under-representation of urbanization in the HadGEM3–PRIMAVERA land cover becomes more apparent when comparing modelled heat-flux characteristics in the London domain (LWC, WIS; Table 1a) with sites in the Shanghai domain (SHA, XIA,
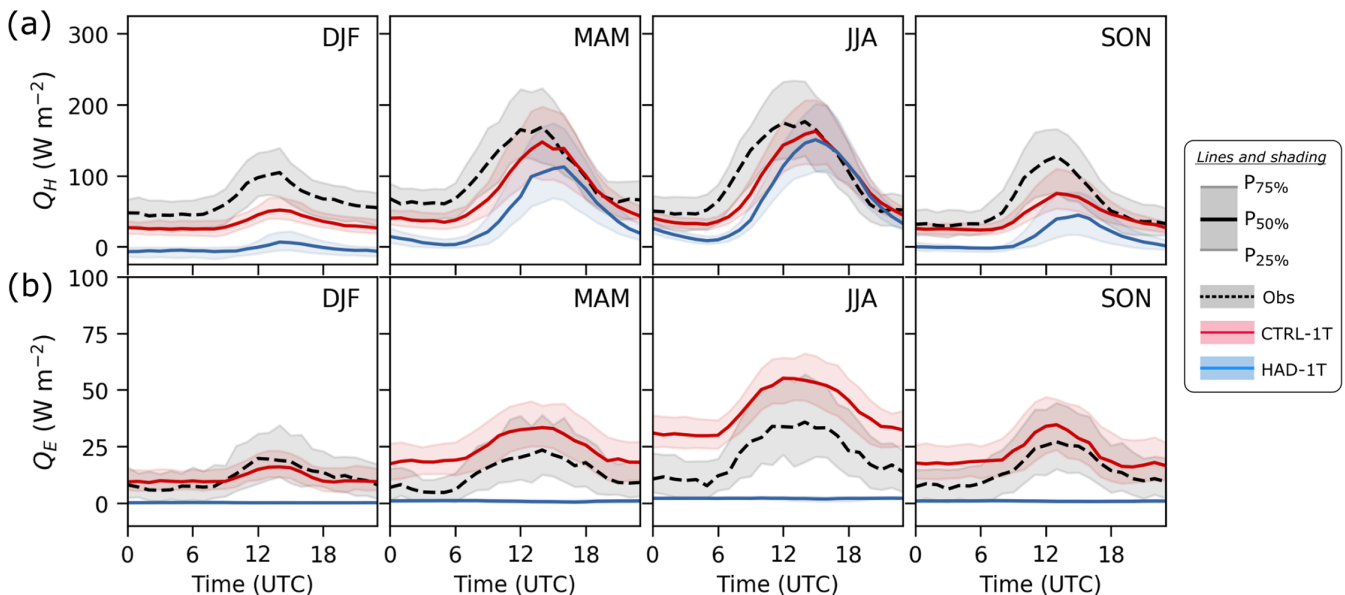
FIGURE 3 Seasonal median diurnal cycles based on hourly fluxes (2011–2013) of turbulent (a) sensible heat flux (Q_H), (b) latent heat flux (Q_E) in central London (KCL site; Appendix A) from observations and offline JULES simulations using the Best (2005) single tile (1T) urban model (GL7.0 configuration) with control (CTRL–1T) and HadGEM3–PRIMAVERA (HAD–1T) parameters and land cover (Table A1; Appendix A) [Colour figure can be viewed at wileyonlinelibrary.com]

LUK; Table 1b). Comparison is undertaken using the monthly ensemble median (2005–2010) surface heat fluxes and EF from the HadGEM3–PRIMAVERA climate simulations for the three ensemble members (Figure 2d-f). To account for radiative forcing differences in the two regions, the daily-mean surface heat fluxes are normalized by the local daily-mean net all-wave radiation (Q_N; Figure 2d,e). The central London LWC site (HadGEM3 f_{Urban} = 97.7%) clearly has the largest 'urban' response, with bulk characteristics similar to the HAD–1T offline test (e.g., negative Q_H occurs in DJF; too low Q_E). The results for the WIS background station are dominated by the large fraction of non-urban surfaces (C3, C4 grass: 66.7%, bare soil: 7.2%), causing much larger Q_E in summer (*cf.* LWC). As the model output for fluxes are diurnal means, the winter Q_N values in the UK are small and often negative (Figure S5a), impacting the sign of ratios (Q_E/Q_N; Q_H/Q_N). The daily-mean DJF for Q_E is mostly positive (Figure S5c) and for Q_H mostly negative (Figure S5b). Similarly, EF (Equation 2) is negative when the daily-mean $Q_H < 0$ W m^{-2} and $|Q_H| > |Q_E|$, resulting in larger variability in the monthly statistics (Figure 2d-f) in winter at the UK sites (*cf.* China).

With the SHA grid-box (Baoshan district of Shanghai; Table 1b) having only 6.3% built land cover but 79.2% vegetation (grass and shrubs), it is unsurprising that the heat fluxes are similar to those modelled at the rural WIS site (UK), especially in MAM and JJA (Figure 2d,e). Similarly, the model fluxes at both LUK (Nanjing/Lukou airport) and XIA (Hangzhou airport) with zero built land cover in HadGEM3 show a typical rural response with the monthly median $EF > 0.5$ year-round. Here, the model assumes 86.3% (XIA) and 90.9% (LUK) vegetated surfaces. As a point of reference, Ao *et al.* (2016a) report observed monthly mean daytime Bowen ratios between 2 and 4.7 (i.e., EF between 0.33 and 0.18) for a central business district in Shanghai (XJH site; Xujiahui district). The modelled median EF at SHA is much higher and ranges between 0.42 (August) and 0.68 (June; Figure 2f). Observed mean daily Q_H peaks in Shanghai can exceed 290 W m^{-2} in the early afternoon in JJA, while Q_E is low (~65 W m^{-2}; Ao *et al.*, 2016a). Misrepresenting the energy partitioning over cities in such a way will negatively impact any use of this data, such as for climate-service applications that use heat-flux ratios to assess urban heat stress for health of citizens or irrigation demands for maintaining green infrastructure or reducing dust resuspension.

4.2 | Urban heat-island intensity

Turbulent heat fluxes (Section 4.1) and surface temperatures play an important role in determining and

co-modulating local boundary-layer dynamics in atmospheric models over cities and therefore impact characteristics of near-surface air temperatures (e.g., Omidvar *et al.*, 2020). Distinct canopy-layer air-temperature differences between urban and rural areas have long been observed worldwide (e.g., literature reviewed in Oke *et al.*, 2017 and Stewart, 2019), particularly a few hours after sunset when the urban heat island is strongest. Radiative and thermal properties of prevalent urban materials, together with the density and volume of buildings, result in heat being effectively stored (ΔQ_S) during the day and released at night when Q_N becomes negative (Equation 1). This, together with Q_F, partially offsets radiative night-time cooling in cities; a process that it is critical to represent in land-surface models.

Figure 4 shows a comparison of seasonal median diurnal T_{air} cycles between observations and climate model (HAD$_{1,2}$; Section 2.3) at LWC (central London) in the UK (Figure 4a) and SHA (Baoshan/Shanghai) in China (Figure 4b). Comparisons of T_{air} at the remaining sites in both domains are shown in Figures S6 and S8.

With nearly 100% built surface cover (Table 1a), the response of the urban Best–1T scheme in JULES dominates the T_{air} characteristics at LWC (Figure 4a). Only in summer (when Q_N is largest) does the model not underestimate the median T_{air}. However, large differences exist in both magnitude and timing of modelled and observed median JJA T_{air} peaks. Compared to the observations, both ensemble members show a ~1 h delay in the morning temperature rise and a further phase shift of the afternoon peaks of 1–2 h, together with a positive offset of the maximum median T_{air} of ~2°C (HAD$_1$) and ~1°C (HAD$_2$). These phase delay features are observed offline in central London (HAD–1T) in Q_H (Figure 3a). The positive afternoon T_{air} bias in JJA at the highly urbanized sites (Figures 4a and S6a) can be partially explained by an over-prediction of surface temperatures (see also JJA $L^{\uparrow}$ bias in Best–1T offline tests in central London; Figure S4b). This feature of the urban model affects the grid-box T_{air} less at sites with lower f_{Urban} (LHR, WIS; Table 1a). While at these sites in JJA HAD$_1$ performs much better (Figure S6b,c), the positive bias persists for HAD$_2$, showing that the initial conditions (Section 2.1) and the resulting response of the atmospheric model in the region have an influence.

During colder seasons, especially winter, there is a persistent negative bias in modelled T_{air} at LWC (Figure 4a). This is stronger at night and in the early morning (up to 2°C in DJF). As discussed in Section 4.1, this is likely related (in part) to the climate model Q_F being 0 W m^{-2}, whereas observations at the highly urbanized LWC are impacted by Q_F. In central London in winter, Q_F can be as large as, or larger than, Q_N and
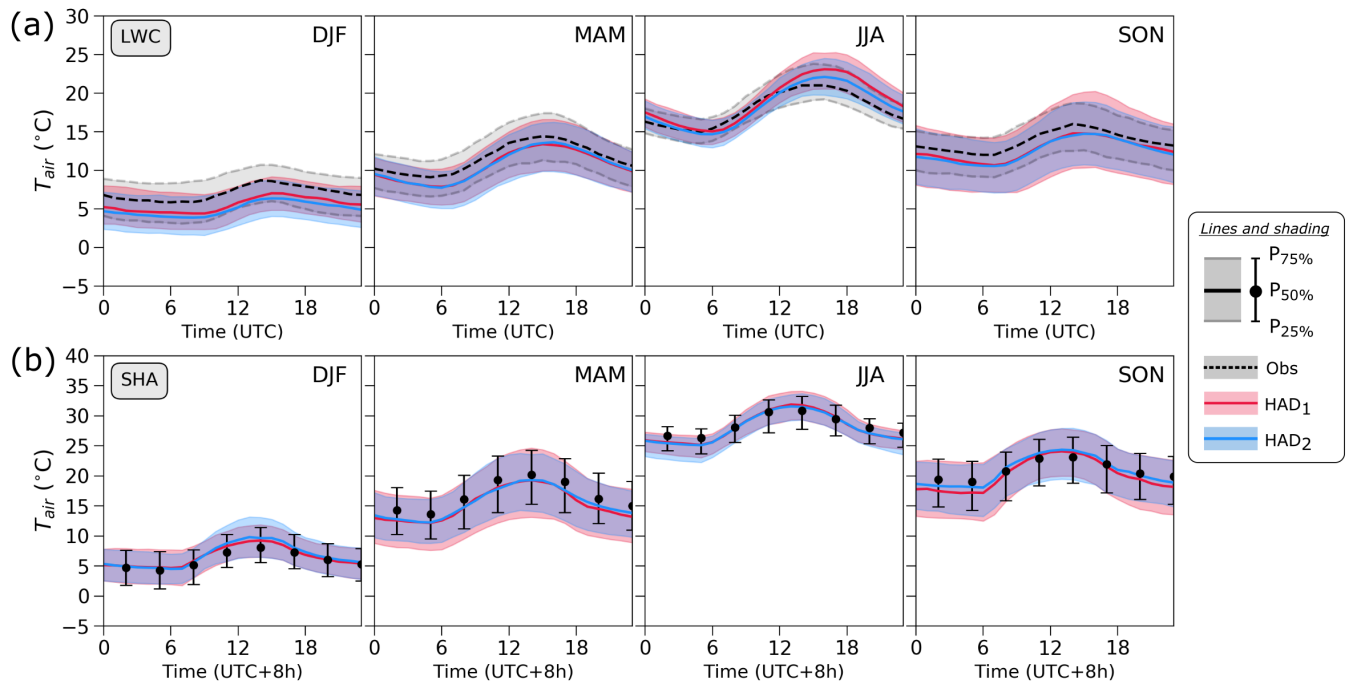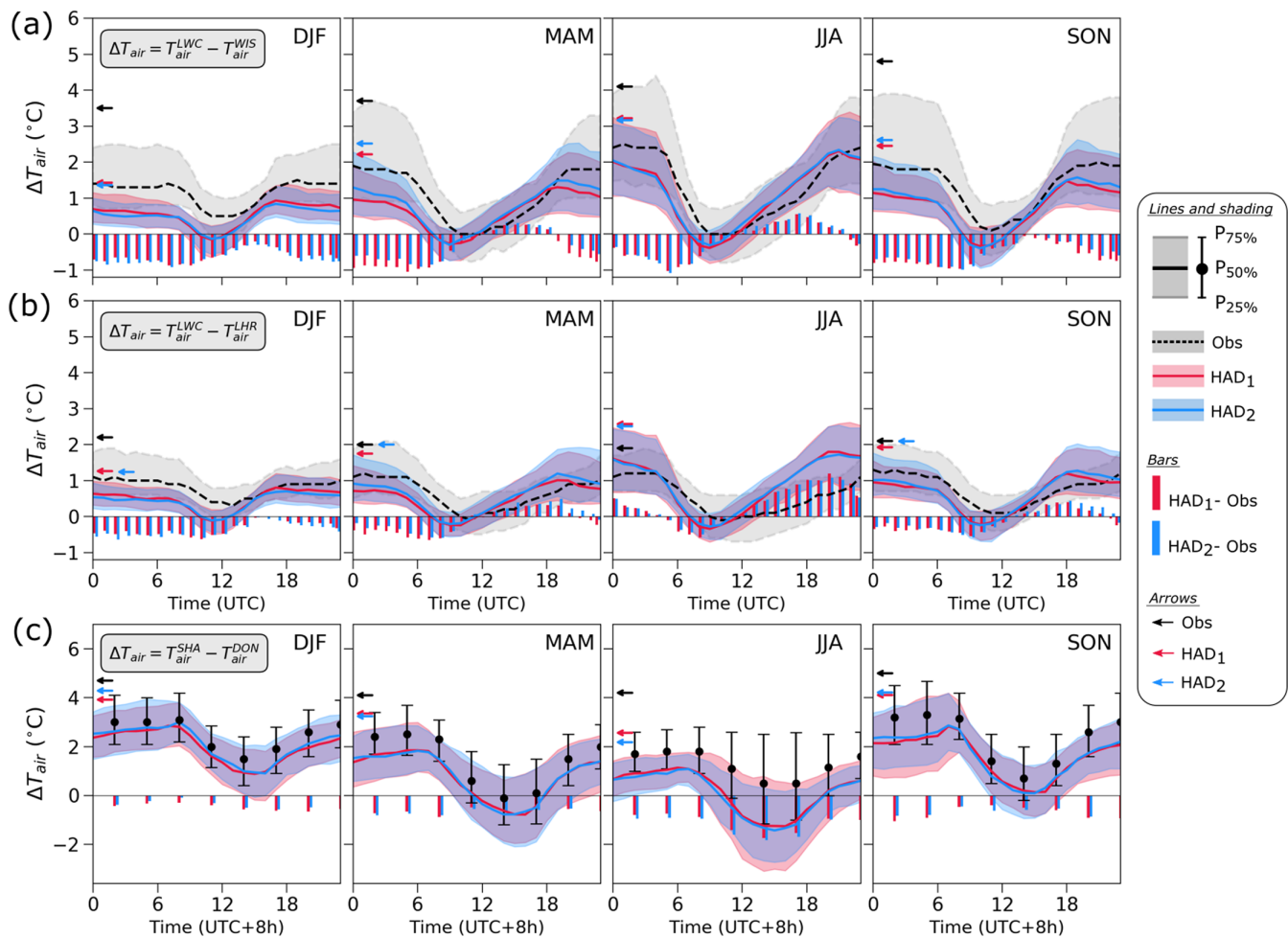
FIGURE 4 HadGEM3–PRIMAVERA (HAD$_{1,2}$) and observed seasonal median diurnal cycles of screen-level air temperature (T_{air}) with inter-quartile range for (a) LWC (London Weather Centre; Table 1a) and (b) SHA (Shanghai/Baoshan; Table 1b). Observations are (a) hourly and (b) 3-hourly. See Figure S6 for SJP, LHR and WIS (UK); Figure S8 for XIA, DON, LIY, LUK (China) [Colour figure can be viewed at wileyonlinelibrary.com]

can therefore be the main driver of the surface energy-balance (Hamilton *et al.*, 2009; Kotthaus and Grimmond, 2012). Offline tests with the JULES Best–1T model in central London (KCL site; Appendix A) show that switching off Q_F in CTRL–1T can account for up to ~1°C difference in DJF T_{air} (Figure S7). This agrees with anthropogenic temperature increments of 0.5–1°C determined by Bohnenstengel *et al.* (2014) for other sites across London. Compared to SJP, located in a highly vegetated park and therefore less impacted by Q_F (*cf.* Figure S2), LWC observes consistently warmer median night-time T_{air} of ~0.6°C year-round (Figure S9), in agreement with Jones and Lister (2009). As both sites are within the same 0.1° model grid-box, these local differences are not represented in the model. Consequently, the comparison of the HadGEM3–PRIMAVERA T_{air} with the statistically cooler SJP site (*cf.* LWC) has a better agreement of median night-time temperatures (except summer), but an exacerbated positive model bias of JJA afternoon temperatures (Figure S6a). These differences in model assessment depending on site choice (within the same model grid-box) show that it is crucial to have measurements available in characteristically urban settings that reflect the added effects of Q_F and storage heat flux from the building volume.

The intensity of the UHI can be assessed by the difference between urban and rural air temperatures.

Figure 5a shows the modelled and observed seasonal median diurnal cycle of screen-level air temperature differences (ΔT_{air}) between LWC and the rural WIS station ~35 km south-west of LWC (Figure 1a,b). Expectedly, the observed median ΔT_{air} is highest (and nearly constant) at night, with largest values in JJA (~2.5°C) and lowest in DJF (~1.5°C). The magnitudes agree with previous long-term observations (Jones and Lister, 2009). In all seasons, the climate model underestimates the median UHI intensity between sunset and sunrise by up to 1°C (DJF), while in MAM and JJA the median ΔT_{air} is overestimated in the afternoon and early evening by up to 0.5°C. Similarly, the large magnitude and seasonal variability of the observed inter-quartile range of ΔT_{air} is not reproduced by the model, indicating a smaller sample spread. Strong differences also exist in the seasonal 90th percentiles of ΔT_{air}, which range between 3.5°C (DJF) to 4.9°C (SON) in the observations and 1.4°C (DJF) and 3.1°C (JJA) in the model ensembles, implying that more extreme urban–rural temperature contrasts are not captured.

Observed T_{air} differences between central London (LWC) and the less urbanized Heathrow airport (LHR; f_{Urban} = 48.3% in GUF$_{2.5\ km}$, 33.1% in HadGEM3; Table 1a) are noticeably smaller compared to LWC–WIS (Figure 5b). The observed median ΔT_{air} between sunset and sunrise is relatively constant throughout the year

FIGURE 5 HadGEM3–PRIMAVERA ($HAD_{1,2}$) and observed seasonal median (lines/markers) diurnal cycles of screen-level air temperature differences (ΔT_{air}) with inter-quartile range (shading/error-bars) for (a) LWC–WIS and (b) LWC–LHR (Table 1a) and (c) SHA–DON (Table 1b), with differences (vertical bars) of the median ΔT_{air} between model ensemble members ($HAD_{1,2}$) and observations (a, b) hourly and (c) 3-hourly; and seasonal 90th percentile of the data (arrows). See Figure S10 for ΔT_{air} for SHA–XIA, SHA–LIY and SHA–LUK (China) [Colour figure can be viewed at wileyonlinelibrary.com]

with only small variations around 1°C, while the model has larger seasonal variability of ΔT_{air} with nocturnal maxima of only ∼0.5°C in DJF, but up to ∼1.8°C in JJA. These magnitudes are very similar to the modelled ΔT_{air} between LWC and the rural WIS site (Figure 5a). While more vegetation surrounds LHR (*cf.* LWC), the site is also affected by anthropogenic heat emissions related to the airport infrastructure and operations.

Systematic differences between $HAD_{1,2}$ and observations also exist at the SHA site (Figure 4b). Nocturnal and early morning median T_{air} are under-estimated in all seasons except winter (DJF), while T_{air} peaks are over-estimated in all seasons except for spring (MAM). This occurs at all sites in the Shanghai domain (Figure S8) and can be clearly traced in the bias patterns of the median 3-hourly T_{air} in each season (Figure 6). As the model f_{Urban} severely under-represents the actual built land-cover fraction at all sites (Figure 1d; Table 1b), the

nocturnal negative model bias tendency partially relates to the lack of urban heat storage/release. For sites in populous cities (e.g., SHA, XIA) with a large building volume, anthropogenic heat emissions are also likely to affect the near-surface air temperatures. While in London and other higher-latitude cities Q_F typically peaks in winter when space heating is needed, subtropical Shanghai and surrounding cities like Hangzhou have Q_F peaks in summer from air-conditioning use. While space cooling is used in both commercial and residential buildings, heating during winter can occur in offices, but is uncommon in residences (Ao *et al.*, 2016a). This could explain the overall better agreement of nocturnal T_{air} in DJF between model and observations, as both sites are situated outside of high-rise central business districts with predominantly commercial buildings and instead are dominated by low-rise buildings (residences and industry).
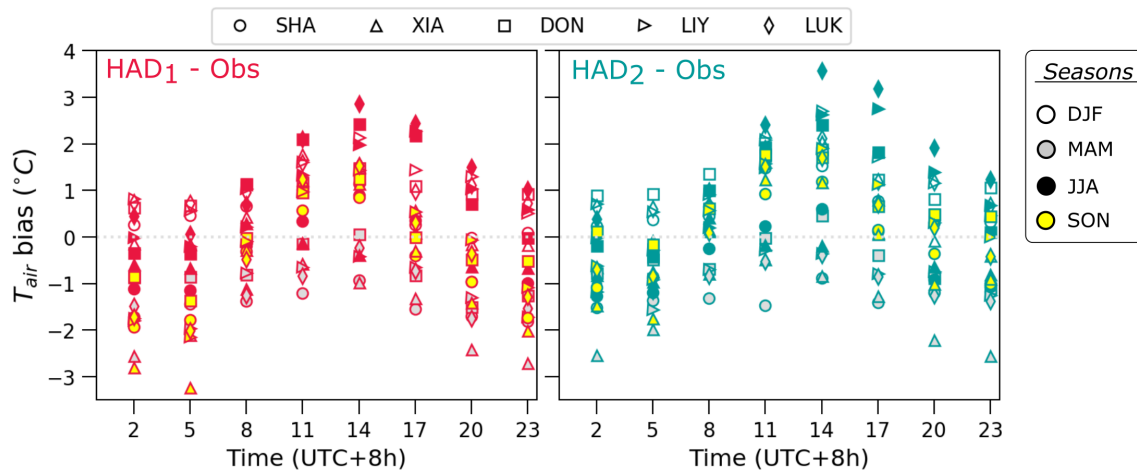
FIGURE 6 Bias (model-observations) of median 3-hourly T_{air} by season at sites in the Shanghai region (Table 1b) for two ensemble members ($HAD_{1,2}$) [Colour figure can be viewed at wileyonlinelibrary.com]

Of the five evaluation sites in the Shanghai region (Table 1b), none qualify as truly rural background stations (Figure S2). However, it may be expected that differences in urbanization levels between SHA, with the highest f_{Urban}, and the other sites are to a degree reflected in ΔT_{air}. However, given the large geographical distances, ΔT_{air} primarily reflects regional climate variations and differences between coastal sites (SHA, XIA), that can be affected by land–sea breeze circulations, and inland stations.

Seasonal diurnal cycles of ΔT_{air} between SHA and DON (Figure 5c; SHA: f_{Urban} = 87.3%, DON: 36.1% based on GUF$_{2.5 km}$; Table 1b) have large magnitudes and variability in both observations and model. As the model f_{Urban} is very low at both sites (SHA: 6.3%; DON: 9.7%), the modelled ΔT_{air} mainly reflects climatological differences. The observed median ΔT_{air} is consistently higher than the modelled equivalent, with particularly large differences at night in MAM and SON and during the day in JJA. This trend agrees with the expected role of urban ΔQ_S and Q_F for SHA. Similar trends are seen in the observed ΔT_{air} between SHA and the other sites in the domain (Figure S10), with the overall smallest differences occurring between the coastal SHA and XIA stations.

At the inland sites (DON, LIY, LUK), a large overestimation of the median JJA screen-level temperatures occurs (Figures 6 and S8). This is smaller for the coastal SHA and XIA stations. Unlike at LWC (Figure 4a), this cannot be attributed to features of the Best–1T model given the largely missing f_{Urban} characterization in the model, but must relate to the regional weather pattern representation by the atmospheric model. Note that differences between surface elevations reported for the stations and the model orography are small.

An increased resolution of global climate models can modify the hydrological cycle simulated, with an increase

in precipitation over land (Vannière *et al.*, 2019). In June and July, the analysis region centred on Shanghai is affected by meso-scale convective systems associated with the passage of the Meiyu front, which results in very intense localized rainfall for short durations (e.g. Guan *et al.*, 2020). While the climate model can resolve the Meiyu frontal system and its passage through the region (A. Volonté, pers. comm.), differences in timing, intensity and location of associated convective precipitation compared to the observations (not shown) may play a role in the JJA temperature bias (Figure 6).

The urban model behaviour (offline and online) in highly urbanized central London suggests the positive JJA model bias of T_{air} in the Shanghai domain will likely be exacerbated once more realistic urban land cover is used. Hence, improving land-cover characteristics is expected to cause model performance to deteriorate at these sites, which prompts the need to further investigate reasons for the bias in the atmospheric model to improve the model performance.

In some seasons, there are large quantitative differences between T_{air} from both ensemble members. Compared to HAD_2, HAD_1 for the UK domain has a larger 75th percentile and sometimes larger median T_{air} (e.g., DJF, JJA; Figures 4a, S6) at all sites across the common period of data availability. Whereas for the China domain, HAD_1 has a lower 25th percentile in MAM and SON and lower night-time medians (Figures 4a, S8). These ensemble differences can be as large as 1–2°C, and hence can have a meaningful impact on potential uses of these data, such as for local climate assessment and planning, if extremes are of interest (Section 4.3). Similarly, the HAD_1 surface heat fluxes in the UK have larger differences to the other ensemble members, especially in summer, while HAD_2 and HAD_3 are more similar (not

shown). As the implications for urban services can be significant, this raises more general questions about the use of probabilistic versus deterministic forecasts.

4.3 | Occurrence of temperature extremes

The frequency of occurrence and magnitude of very warm (e.g., heat waves) or cold air temperatures in cities can be affected by the UHI (e.g., Lemonsu *et al.*, 2015; Ramamurthy and Bou-Zeid, 2017; Ao *et al.*, 2019) and can be further exacerbated by feedback mechanisms with Q_F (e.g., through air-conditioning; Takane *et al.*, 2020). Representing such processes in climate projections for cities is crucial for various urban applications, such as related to thermal comfort/heat stress and associated mitigation strategies (e.g., urban greening; Zölch *et al.*, 2016).

While the HadGEM3–PRIMAVERA simulation period is too short to derive statistically robust occurrence likelihoods of heat waves, the occurrence frequency of recent temperature extremes as represented in the model and captured in the observations can be compared for the available output periods. Modelled and observed seasonal frequency distributions (normalized) of T_{air} are compared for daytime and night-time periods in the UK and China domains (Figure 7; Figures S11, S12). For the UK, the analysis of hourly data varies daylength between seasons (all UTC): DJF 09:00–16:00; MAM 07:00–20:00; JJA 06:00–21:00; and SON 08:00–18:00. In the Shanghai domain, daytime length varies much less through the year and the observation frequency is only 3-hourly. Hence, a constant daytime period (06:00–18:00 UTC + 8 h) is used. Common periods and data frequencies for both the observations and model output are used (see Section 2.3).

The overall shape of the observed T_{air} frequency distributions during day and night is captured by the model in both domains, with a mostly good agreement for the range of values and skewness patterns (Figure 7). However, some of the discrepancies between the diurnal air temperatures ranges (identified in Section 4.2) are reflected in the tails of the distributions. The higher observed DJF air temperatures at the London city-centre site LWC (Figure 4a) are reflected in the cool-end tail of the T_{air} distributions (Figure 7a). Similarly, the overprediction of JJA and SON daytime temperatures by the model impact the warm-end T_{air} tail (stronger for HAD$_2$). At the rural WIS site (Figure 7b), minimum nocturnal temperatures are over-predicted throughout the year, in agreement with characteristics of the 25th percentiles of the seasonal diurnal cycles (Figure S6c), likely linked to the higher model f_{Urban} (Table 1a).

At the SHA (Shanghai/Baoshan) site in the China domain (Figure 7c), the observed nocturnal maximum T_{air} in MAM and JJA are slightly warmer than in the model (see also Figure 4b). This could be connected to the limited nocturnal heat release from building volumes, with the low urban tile weighting in the grid-box (Table 1b). The largest difference between the T_{air} tails in the Shanghai domain is found in the JJA daytime maximum temperatures at the inland sites (Figure 7d for DON; Figure S12b,c for LIY, LUK), with a strong overprediction (much longer tails) from both ensemble members. This agrees with earlier discussions (Section 4.2). Nocturnal distributions of temperatures at these sites are much better modelled in all seasons, but as for the coastal SHA and XIA sites the warm-end tail is slightly underpredicted in MAM and JJA.

The frequency of days with extreme T_{air} for each simulation year is analysed (Figures 8, 9) using observed and modelled daily maximum and minimum temperatures at each site. Cold and hot thresholds are set to be below/above the 25th/75th percentiles of observed temperatures (DJF, JJA, respectively) at all sites. For both domains, the same two cold thresholds are used: 0 and 5°C. However, the hot extreme differs between domains, with London thresholds (25/30°C) lower than for the Shanghai region (30/35°C). A day is counted as cold (hot) if T_{air} over 24 h is detected below (above) these limits. Qualitatively, the results (Figures 8 and 9) are insensitive to the exact threshold chosen (as long as they are within the distribution tails). The frequency is relative to the total number of days in the year that have valid data, considering both missing observations and model output. In the UK, hourly T_{air} data are analysed, whereas in China 3-hourly samples (Section 3) of the hourly model output are used to match the observations (Section 2.3).

As the two central London sites (LWC, SJP; Figure 8) are within the same model grid-box, the observed frequencies of hot and cold days reflect the climatological differences between the stations (Section 4.2). The park station (SJP) has a notably higher occurrence of cold days *cf.* LWC with its more extensive built/impervious surfaces and slower nocturnal radiative cooling (Figure S9). The observed relative occurrence of hot days is only slightly higher at LWC (*cf.* SJP), in agreement with the expected smaller role of UHI intensity type processes during the day.

The modelled cold days for HAD$_{1,2}$ agree better with SJP (*cf.* LWC). In 2007 and 2008, the modelled occurrence of both $T_{air} < 5°C$ and $T_{air} < 0°C$ at LWC are too high. Whereas for rural WIS the model ensembles in all years underestimate the occurrence frequency of $T_{air} < 0°C$. This agrees with overestimation of the 25th percentile of T_{air} in DJF (*cf.* observations, Figure S6c).
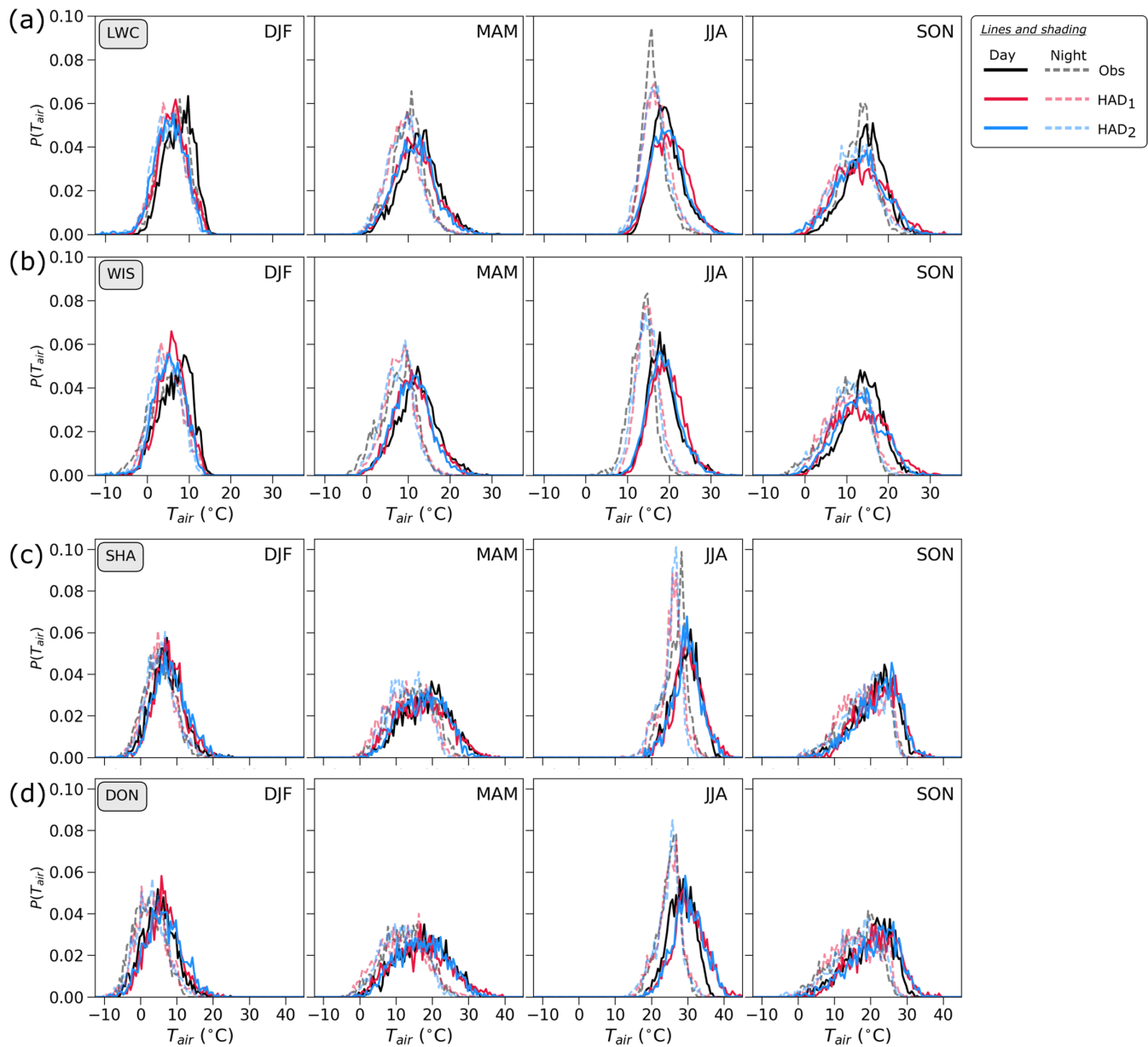
FIGURE 7 HadGEM3–PRIMAVERA (HAD$_{1,2}$) and observed normalized frequency distributions of T_{air} (0.5°C bin size) per season from daytime and night-time samples for (a) LWC (London Weather Centre), (b) WIS (Wisley), (c) SHA (Shanghai/Baoshan) and (d) DON (Dongtai). Samples in (a,b) are hourly, in (c,d) 3-hourly with model output frequency reduced to match the observations. See Figure S11 for results at SJP and LHR (UK); Figure S12 for XIA, LIY and LUK (China) [Colour figure can be viewed at wileyonlinelibrary.com]

This may be partly caused by f_{Urban} being too high in the model relative to the actual surroundings of the site (26.1% and a maximum of 75.5% in a 3-by-3 grid-box area *versus* 3.5% and 23.1% in GUF$_{2.5\ km}$ and GUF$_{25\ km}$; Table 1a).

In the China domain, model performance differs between the coastal, highly urbanized sites (SHA, XIA) and the sites inland and to the north (DON, LIY, LUK). This is reflected in the temperature extremes (Figure 9). For the latter, in all years, the model noticeably over-predicts the occurrence of hot ($T_{air} > 30°C$) and very hot

($T_{air} > 35°C$) days compared to the observations, in agreement with the JJA bias patterns (Figure 6). Cold days at these sites are better predicted, with no clear inter-annual trend of the model performance. At SHA and XIA, hot days are over-predicted in some years, while in others (2005, 2007 at XIA) the occurrence frequency is under-predicted. Both the model and the observations have a lower occurrence of cold days at the more coastal stations. However, the model results do not reflect the typical urban response of the land-surface at these sites due to the severe under-representation of urban land
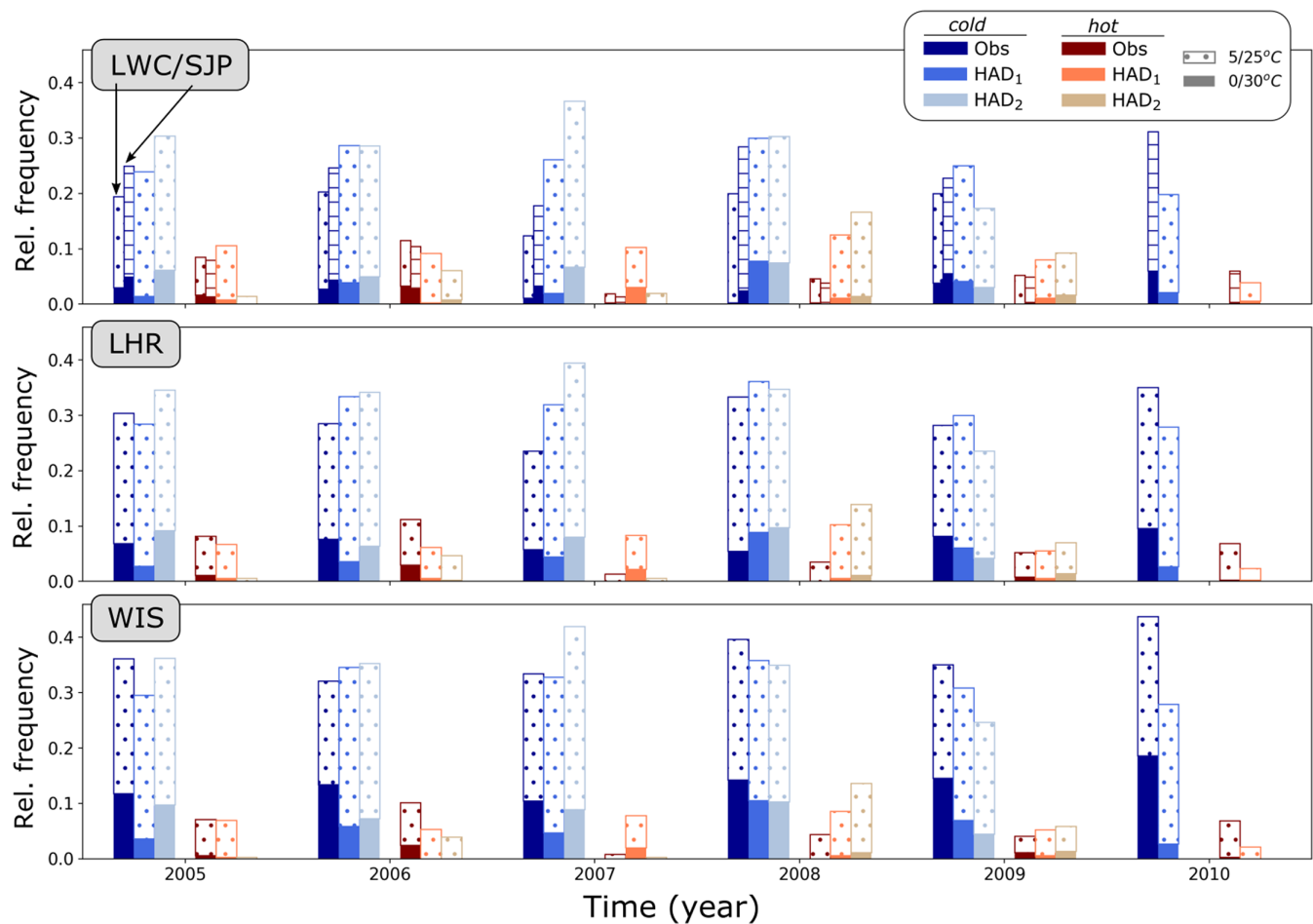
FIGURE 8 Relative frequency of days within a year for which cold/hot screen-level temperatures (T_{air}) were observed and modelled (HAD$_{1,2}$) at sites in the south-east UK domain. The two cold/warm thresholds used for T_{air} are (i) 5/25°C (hatched bars) and (ii) 0/30°C (filled bars). Note there is an absence of data for 2010 for HAD$_2$ and LWC observations. As LWC and SJP are within the same model grid-box they are shown together (LWC: dotted bars; SJP: hatched) [Colour figure can be viewed at wileyonlinelibrary.com]

cover (Secions 2.2 and 3). Hence, some of the better agreement observed here may deteriorate if more realistic land cover is used.

5 | CONCLUSIONS

High-resolution (~10 km; N1280) global climate simulations (2005–2010) with the Met Office HadGEM3 model are analysed over large urban areas in the south-east UK (London) and south-east China (Shanghai, Hangzhou, Nanjing region) to study the response of modelled surface heat fluxes and diagnostic screen-level temperatures (T_{air}) to urbanization levels. Modelled T_{air} is evaluated using weather station data. The climate model uses a simple urban slab scheme with prescribed, globally fixed parameters (JULES–GL7.0) and land cover derived from IGBP.

While any detected T_{air} bias could be partially attributed to bias in the large-scale atmospheric model (and

that needs to be investigated further), differences can also be linked to both the model land cover and the specifications used in the urban land-surface model. We draw the following conclusions regarding the key factors affecting representation of urban signals in the simulations and the potential for improvements if model output is intended to inform applications in urban areas (e.g., urban climate services):

- The *representation of urban land cover* is identified as the primary source of bias in the JULES land-surface model. Comparisons of recent (2011) high-resolution reference data (GUF) to the IGBP-based model land cover in China, suggest these severely underestimate both the spatial extent and magnitudes of urbanization. China's most populous megacity, Shanghai, in the model only covers four ~10 km grid-boxes with a maximum f_{Urban} of 60.2%, while other conurbations are not captured at all. Hence, the modelled response of surface
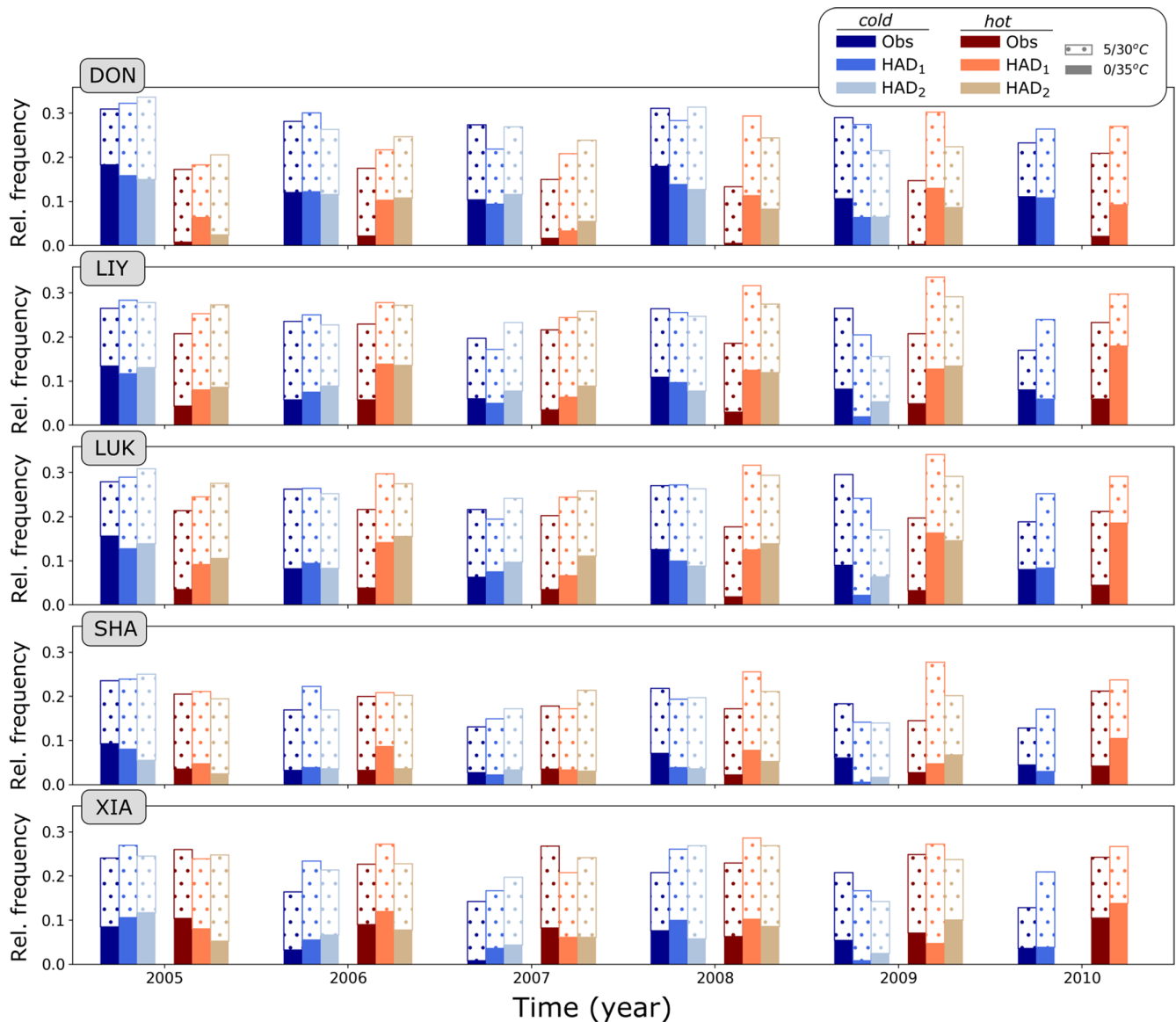
FIGURE 9 As Figure 8, but for sites in the south-east China domain with two cold/hot thresholds: (i) 5/30°C (hatched bars) and (ii) 0/35°C (filled bars). Note that HAD$_2$ has no model output for 2010 [Colour figure can be viewed at wileyonlinelibrary.com]

heat fluxes that drive boundary-layer dynamics are predominantly rural/non-urban in nature. This is inevitably reflected in the near-surface air temperatures, as urban (e.g., heat-island) effects are absent. These, for example, can impact heat-wave intensity climatologies. For some sites analysed in the Shanghai region, it is anticipated that resolving the land-cover characterization may increase the JJA T_{air} bias, prompting the need to further investigate reasons for the bias in other parts of the model system. The land cover of Greater London is more realistic in the model, but non-built surface covers (vegetation, water) are too low in central London (model $f_{Urban} \approx 100\%$) causing bias in the energy partitioning (e.g., Q_E too small). Thus, it is concluded both too small and too large f_{Urban} negatively impact model

results and therefore, the use of these data for applications such as urban climate services (e.g., heat stress assessment or external water use requirements). Hence, use of appropriate (current and future) urban land-cover information is crucial, and perhaps the part of the modelling chain that is easiest to fix as high-resolution satellite-derived global land-cover products have become more widely available in recent years (e.g., high-resolution (~300 m) global land-cover data of the European Space Agency's Climate Change Initiative, ESA-CCI). High-resolution climate simulations in future climates need to include potential future changes in land cover, such as urban expansion and/or changes in land-cover types (e.g., from enhanced green infrastructure; e.g., Li *et al.*, 2017, Carter, 2018).

- *Urban anthropogenic emissions of heat and water* are absent in the current simulations. This causes biases in turbulent heat fluxes and T_{air}. In both London and Shanghai, Q_F plays an important role in the surface energy balance. Currently, JULES has the capability to prescribe Q_F to the urban tile as spatially unvarying monthly values. Offline tests of the HadGEM3–PRIMAVERA configuration in central London showed that even this simplistic representation can improve the model performance if suitable magnitudes of Q_F are used. As retrieval of local-scale energy-consumption information is challenging at the global scale, Q_F should ideally be modelled based on more accessible parameters like population density and temperature-dependent heating/cooling demands (e.g., Sailor and Vasireddy, 2006; Lindberg *et al.*, 2013). Anthropogenic water emissions also can have a non-negligible effect on the urban energy balance in some cities (e.g. street cleaning, irrigation; Ao *et al.*, 2018, Dou *et al.*, 2019), but currently only vegetated tiles of JULES can be irrigated.

- *Urban scheme physics and parameters* used in HadGEM3–PRIMAVERA (JULES–GL7.0) have fixed singular parameter values, making appropriate choices to represent cities worldwide challenging. For example, the selected default urban albedo (0.18) is large compared to central London ($\sim$0.11 from observations; Kotthaus and Grimmond, 2014a) and central Shanghai ($\sim$0.14; Ao *et al.*, 2016b). This reduces the energy input into the urban system (Q_N), which, combined with the absence of Q_F in the model, can cause underprediction of Q_H. In London, the large thermal inertia of the Best–1 T urban scheme causes a $\sim$ 1-h delay of temperature increase in the morning and up to 2-h delay in the afternoon T_{air} peak in the climate model output. Furthermore, modelled JJA daytime temperatures in central London are overestimated by up to 2°C. This is partially explained by the large heat capacity and roughness length for heat used in the scheme. The JULES two-tile canopy model MORUSES, with separate surface-energy balance calculations for roofs and street canyons, can improve this by explicitly modelling the bulk radiative, thermal and aerodynamic parameters as a function of building morphology (Hertwig *et al.*, 2020). At grid-box scale, the fast response of the (insulated) roof tile to radiative forcing can partially offset the large heat storage and correspondingly delayed sensible heat flux of the canyon tile (Porson *et al.*, 2010). However, urban canopy models require more characteristics than built land-cover fractions, which are currently not available globally; for example, roof and street fractions need to be separated and mean building heights, height-to-width ratios of street canyons and radiative/thermal material

characteristic of built surfaces need to be known. It is also noted that the impact of very tall buildings is not captured at all in the JULES urban schemes. In central Shanghai, for example, over 1200 buildings are taller than 100 m and extend to 632 m (Tan *et al.*, 2015). Hence, the buildings are much larger than the local topography (mean elevation above sea level is 4 m). Given the increasing verticality of cities worldwide, more research is needed to better understand the impacts of tall buildings on the urban surface-energy balance and how these can be represented in urban models (Barlow *et al.*, 2017).

ORCID

Denise Hertwig https://orcid.org/0000-0002-2483-2675
Sue Grimmond https://orcid.org/0000-0002-3166-9415
Pier Luigi Vidale https://orcid.org/0000-0002-1800-8460
Patrick C. McGuire https://orcid.org/0000-0001-6592-4966

REFERENCES

Allen L, Lindberg F, Grimmond CSB. 2011. Global to city scale urban anthropogenic heat flux: model and variability. *International Journal of Climatology* 31: 1990–2005. https://doi.org/10.1002/joc.2210.

Ao X, Grimmond CS, Ward HC, Gabey AM, Tan J, Yang X, Liu D, Zhi X, Liu H, Zhang N. 2018. Evaluation of the Surface Urban Energy and Water Balance Scheme (SUEWS) at a dense urban site in Shanghai: sensitivity to anthropogenic heat and irrigation. *Journal of Hydrometeorology* 19: 1983–2005. https://doi.org/10.1175/JHM-D-18-0057.1.

Ao X, Grimmond CSB, Chang Y, Liu D, Tang Y, Hu P, Wang Y, Zou J, Tan J. 2016a. Heat, water and carbon exchanges in the tall megacity of Shanghai: challenges and results. *International Journal of Climatology* 36: 4608–4624. https://doi.org/10.1002/joc.4657.

Ao X, Grimmond CSB, Liu D, Han Z, Hu P, Wang Y, Zhen X, Tan J. 2016b. Radiation fluxes in a Business District of Shanghai, China. *Journal of Applied Meteorology and Climatology* 55: 2451–2468. https://doi.org/10.1175/JAMC-D-16-0082.1.

Ao X, Wang L, Zhi X, Gu W, Yang H, Li D. 2019. Observed synergies between urban heat islands and heat waves and their controlling factors in Shanghai, China. *Journal of Applied Meteorology and Climatology* 58: 1955–1972. https://doi.org/10.1175/JAMC-D-19-0073.1.

Baklanov A, Grimmond CSB, Carlson D, Terblanche D, Tang X, Bouchet V, Lee B, Langendijk G, Kolli R, Hovsepyan A. 2018. From urban meteorology, climate and environment research to integrated city services. *Urban Climate* 23: 330–341. https://doi.org/10.1016/j.uclim.2017.05.004.

Barlow J, Best M, Bohnenstengel SI, Clark P, Grimmond S, Lean H, Co-authors. 2017. Developing a research strategy to better understand, observe and simulate urban atmospheric processes at kilometre to sub-kilometre scales. *Bulletin of the American Meteorological Society* 98: ES261–ES264. https://doi.org/10.1175/BAMS-D-17-0106.1.

Best MJ. 2005. Representing urban areas within operational numerical weather prediction models. *Boundary-Layer Meteorology* 114: 91–109. https://doi.org/10.1007/s10546-004-4834-5.

Best MJ, Grimmond CSB. 2016a. Modeling the partitioning of turbulent fluxes at urban sites with varying vegetation cover. *Journal of Hydrometeorology* 17(10): 2537–2553. https://doi.org/10.1175/JHM-D-15-0126.1.

Best MJ, Grimmond CSB. 2016b. Modelling the Bowen ratio at a number of urban sites over a range of vegetation cover. *Journal of Hydrometeorology* 17: 2537–2553. https://doi.org/10.1175/JHM-D-15-0126.1.

Best MJ, Grimmond CSB, Villani MG. 2006. Evaluation of the urban tile in MOSES using surface energy balance observations. *Boundary-Layer Meteorology* 118: 503–525. https://doi.org/10.1007/s10546-005-9025-5.

Best MJ, Pryor M, Clark DB, Rooney GG, Essery RLH, Co-authors. 2011. The Joint UK Land Environment Simulator (JULES), model description part 1: energy and water fluxes. *Geoscientific Model Development* 4(3): 677–699. https://doi.org/10.5194/gmd-4-677-2011.

Bohnenstengel SI, Evans S, Clark PA, Belcher S. 2011. Simulations of the London urban heat Island. *Quarterly Journal of the Royal Meteorological Society* 137: 1625–1640. https://doi.org/10.1002/qj.855.

Bohnenstengel SI, Hamilton I, Davies M, Belcher SE. 2014. Impact of anthropogenic heat emissions on London's temperatures. *Quarterly Journal of the Royal Meteorological Society* 140: 687–698. https://doi.org/10.1002/qj.2144.

Carter JG, 2018. Urban climate change adaptation: Exploring the implications of future land cover scenarios. *Cities* 77: 73–80. https://doi.org/10.1016/j.cities.2018.01.014

Clark DB, Mercado LM, Sitch S, Jones CD, Gedney N et al. 2011. The Joint UK Land Environment Simulator (JULES), model description part 2: carbon fluxes and vegetation dynamics. *Geoscientific Model Development* 4(3): 701–722. https://doi.org/10.5194/gmd-4-701-2011.

Cortekar J, Bender S, Brune M, Groth M. 2016. Why climate change adaptation in cities needs customised and flexible climate services. *Climate Services* 4: 42–51. https://doi.org/10.1016/j.cliser.2016.11.002.

Cox PM. 2001. *Description of the TRIFFID dynamic global vegetation model*. Technical Note. Hadley Centre, Met Office: Bracknell, UK, 24.

Cui L, Shi J. 2012. Urbanization and its environmental effects in Shanghai, China. *Urban Climate* 2: 1–15. https://doi.org/10.1016/j.uclim.2012.10.008.

Daniel M, Lemonsu A, Deque M, Somot S, Alias A, Masson V. 2019. Benefits of explicit urban parameterization in regional climate modeling to study climate and city interactions. *Climate Dynamics* 52: 2745–2764. https://doi.org/10.1007/s00382-018-4289-x.

Dou J, Grimmond S, Cheng Z, Miao S, Feng D, Liao M. 2019. Summertime surface energy balance fluxes at two Beijing sites. *International Journal of Climatology* 39: 2793–2810. https://doi.org/10.1002/joc.5989.

Esch T, Bachofer F, Heldens W, Hirner A, Marconcini M et al. 2018. Where we live—a summary of the achievements and planned evolution of the global urban footprint. *Remote Sensing* 10(6), 895. https://doi.org/10.3390/rs10060895

Esch T, Heldens W, Hirner A, Keil M, Marconcini M, Co-authors. 2017. Breaking new ground in mapping human settlements from space – the global urban footprint. *ISPRS Journal of Photogrammetry and Remote Sensing* 134: 30–42. https://doi.org/10.1016/j.isprsjprs.2017.10.012.

Esch T, Schenk A, Ullmann T, Thiel M, Roth A, Dech S. 2011. Characterization of land cover types in TerraSAR-X images by combined analysis of speckle statistics and intensity information. *IEEE Transactions on Geoscience and Remote Sensing* 49 (6): 1911–1925. https://doi.org/10.1109/TGRS.2010.2091644.

Essery R, Best M, Cox P, 2001. MOSES 2.2 Technical Documentation. Met Office Hadley Centre, Hadley Centre Technical Note 30

Gabey AM, Grimmond CSB, Capel-Timms I. 2019. Anthropogenic heat flux: advisable spatial resolutions when input data are scarce. *Theoretical and Applied Climatology* 135(1): 791–807. https://doi.org/10.1007/s00704-018-2367-y.

Goldbach A, Kuttler W. 2013. Quantification of turbulent heat fluxes for adaptation strategies within urban planning. *International Journal of Climatology* 33: 143–159. https://doi.org/10.1002/joc.3437.

Grimmond CS, Oke TR. 1999. Aerodynamic properties of urban areas derived from analysis of surface form. *Journal of Applied Meteorology* 38: 1262–1292. https://doi.org/10.1175/1520-0450(1999)038<1262:APOUAD>2.0.CO;2.

Grimmond CS, Oke TR. 2002. Turbulent heat fluxes in urban areas: observations and a Local-scale Urban Meteorological Parameterization Scheme (LUMPS). *Journal of Applied Meteorology* 41: 792–810. https://doi.org/10.1175/1520-0450(2002)041<0792:THFIUA>2.0.CO;2.

Grimmond CSB, Best M, Barlow J, Arnfield AJ, Baik JJ, Baklanov A, Belcher S, Bruse M, Calmet I, Chen F, Clark P, Dandou A, Erell E, Fortuniak K, Hamdi R, Kanda M, Kawai T, Kondo H, Krayenhoff S, Lee SH, Limor SB, Martilli A, Masson V, Miao S, Mills G, Moriwaki R, Oleson K, Porson A,

Sievers U, Tombrou M, Voogt J, Williamson T. 2009. *Urban Surface Energy Balance Models: Model Characteristics and Methodology for a Comparison Study*. Springer Berlin Heidelberg: Berlin Heidelberg, 97–123. https://doi.org/10.1007/978-3-642-00298-4_11.

Grimmond S. 2007. Urbanization and global environmental change: local effects of urban warming. *Geographical Journal* 173: 83–88. https://doi.org/10.1111/j.1475-4959.2007.232_3.x.

Grimmond S, Bouchet V, Molina LT, Baklanov A, Tan J, Schlünzen KH, Mills G, Golding B, Masson V, Ren C, Voogt J, Miao S, Lean H, Heusinkveld B, Hovespyan A, Teruggi G, Parrish P, Joe P. 2020. Integrated urban hydrometeorological, climate and environmental services: concept, methodology and key messages. *Urban Climate* 33: 2212–0955. https://doi.org/10.1016/j.uclim.2020.100623.

Guan P, Chen G, Zeng W, Liu Q. 2020. Corridors of Mei-Yu-season rainfall over eastern China. *Journal of Climate* 33: 2603–2626. https://doi.org/10.1175/JCLI-D-19-0649.1.

Hamilton I, Davies M, Steadman P, Stone A, Ridley I, Evans S. 2009. The significance of the anthropogenic heat emissions of London's buildings: A comparison against captured shortwave solar radiation. *Building and Environment* 44: 807–817. https://doi.org/10.1016/j.buildenv.2008.05.024.

Hertwig D, Grimmond S, Hendry MA, Saunders B, Wang Z, Jeoffrion M, Vidale PL, McGuire PC, Bohnenstengel SI, Ward HC, Kotthaus S. 2020. Urban signal in high-resolution weather and climate simulations: the role of land-surface characterisation. *Theoretical and Applied Climatology* 142: 701–728. https://doi.org/10.1007/s00704-020-03294-1.

Hewitt HT, Copsey D, Culverwell ID, Harris CM, Hill RSR, Co-authors. 2011. Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system. *Geoscientific Model Development* 4: 223–253. https://doi.org/10.5194/gmd-4-223-2011.

Iamarino M, Beevers S, Grimmond CSB. 2012. High-resolution (space, time) anthropogenic heat emissions: London 1970–2025. *International Journal of Climatology* 32: 1754–1767. https://doi.org/10.1002/joc.2390.

Jones PD, Lister DH. 2009. The urban heat Island in Central London and urban-related warming trends in Central London since 1900. *Weather* 64: 323–327. https://doi.org/10.1002/wea.432.

Kent CW, Grimmond S, Gatey D, Hirano K. 2019. Urban morphology parameters from global digital elevation models: implications for aerodynamic roughness for wind-speed estimation. *Remote Sensing of Environment* 221: 316–339. https://doi.org/10.1016/j.rse.2018.09.024.

Kotthaus S, Grimmond CSB. 2012. Identification of micro-scale anthropogenic CO_2, heat and moisture sources—processing eddy covariance fluxes for a dense urban environment. *Atmospheric Environment* 57: 301–316. https://doi.org/10.1016/j.atmosenv.2012.04.024.

Kotthaus S, Grimmond CSB. 2014a. Energy exchange in a dense urban environment part I: temporal variability of long-term observations in Central London. *Urban Climate* 10: 261–280. https://doi.org/10.1016/j.uclim.2013.10.002.

Kotthaus S, Grimmond CSB. 2014b. Energy exchange in a dense urban environment part II: impact of spatial heterogeneity of the surface. *Urban Climate* 10: 281–307. https://doi.org/10.1016/j.uclim.2013.10.001.

Landauer M, Juhola S, Klein J. 2019. The role of scale in integrating climate change adaptation and mitigation in cities. *Journal of Environmental Planning and Management* 62(5): 741–765. https://doi.org/10.1080/09640568.2018.1430022.

Lean H, Bornemann FJ, Tang Y, Carter E, 2011. Experiences with a 1.5km version of the Met Office Unified Model for short range forecasting. 24th Conference on Weather and Forecasting/20th Conference on Numerical Weather Prediction, Seattle, WA, American Meteorological Society, https://ams.confex.com/ams/91Annual/webprogram/Paper177409.html

Lemonsu A, Viguié V, Daniel M, Masson V, 2015. Vulnerability to heat waves: impact of urban expansion scenarios on urban heat Island and heat stress in Paris (France), *Urban Climate* 14, 586–605. https://doi.org/10.1016/j.uclim.2015.10.007.

Li X, Chen G, Liu X, Liang X, Wang S, Chen Y, Pei F, Xu X. 2017. A new global land-use and land-cover change product at a 1-km resolution for 2010 to 2100 based on human–environment interactions. *Annals of the American Association of Geographers* 107: 1040–1059. https://doi.org/10.1080/24694452.2017.1303357.

Lindberg F, Grimmond CSB, Yogeswaran N, Kotthaus S, Allen L. 2013. Impact of city changes and weather on anthropogenic heat flux in Europe 1995–2015. *Urban Climate* 4: 1–15. https://doi.org/10.1016/j.uclim.2013.03.002.

Loridan T, Grimmond CSB. 2012a. Characterization of energy flux partitioning in urban environments: links with surface seasonal properties. *Journal of Applied Meteorology and Climatology* 51: 219–241. https://doi.org/10.1175/JAMC-D-11-038.1.

Loridan T, Grimmond CSB. 2012b. Multi-site evaluation of an urban land-surface model: intra-urban heterogeneity, seasonality and parameter complexity requirements. *Quarterly Journal of the Royal Meteorological Society* 138: 1094–1113. https://doi.org/10.1002/qj.963.

Loveland TR, Reed BC, Brown JF, Ohlen DO, Zhu Z, Yang L, Merchant JW. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing* 21: 1303–1330. https://doi.org/10.1080/014311600210191.

Macdonald R, Griffiths R, Hall D. 1998. An improved method for the estimation of surface roughness of obstacle arrays. *Atmospheric Environment* 32: 1857–1864. https://doi.org/10.1016/S1352-310(97)00403-2.

Met Office. 2006. *MIDAS: UK Hourly Weather Observation Data*. NCAS British Atmospheric Data Centre http://catalogue.ceda.ac.uk/uuid/916ac4bbc46f7685ae9a5e10451bae7c, Accessed: 1st January 2020.

Mi Z, Guan D, Liu Z, Liu J, Viguié V, Fromer N, Wang Y. 2019. Cities: the core of climate change mitigation. *Journal of Cleaner Production* 207: 582–589. https://doi.org/10.1016/j.jclepro.2018.10.034.

Oke TR, 1973. City size and the urban heat island. Atmospheric Environment (1967) 7, 769–779, DOI: https://doi.org/10.1016/0004-6981(73)90140-6

Oke TR, Mills G, Christen A, Voogt JA. 2017. *Urban Climates*. Cambridge University Press: Cambridge, MA. https://doi.org/10.1017/9781139016476.

Oleson KW, Bonan GB, Feddema J, Vertenstein M. 2008b. An urban parameterization for a global climate model. Part II: sensitivity to input parameters and the simulated urban heat island in offline simulations. *Journal of Applied Meteorology and Climatology* 47: 1061–1076. https://doi.org/10.1175/2007JAMC1598.1.

Oleson KW, Bonan GB, Feddema J, Vertenstein M, Grimmond CS. 2008a. An urban parameterization for a global climate model. Part I: formulation and evaluation for two cities. *Journal of Applied Meteorology and Climatology* 47: 1038–1060. https://doi.org/10.1175/2007JAMC1597.1.

Omidvar H, Bou-Zeid E, Li Q, Mellado J, Klein P. 2020. Plume or bubble? Mixed-convection flow regimes and city-scale circulations. *Journal of Fluid Mechanics* 897: A5. https://doi.org/10.1017/jfm.2020.360.

Porson A, Clark PA, Harman IN, Best MJ, Belcher SE. 2010. Implementation of a new urban energy budget scheme in the MetUM. Part I: description and idealized simulations. *Quarterly Journal of the Royal Meteorological Society* 136: 1514–1529. https://doi.org/10.1002/qj.668.

Ramamurthy P, Bou-Zeid E. 2017. Heatwaves and urban heat islands: a comparative analysis of multiple cities. *Journal of Geophysical Research–Atmospheres* 122: 168–178. https://doi.org/10.1002/2016JD025357.

Roberts MJ, Baker A, Blockley EW, Calvert D, Coward A, Hewitt HT, Jackson LC, Kuhlbrodt T, Mathiot P, Roberts CD, Schiemann R, Seddon J, Vannière B, Vidale PL. 2019. Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. *Geoscientific Model Development* 12: 4999–5028. https://doi.org/10.5194/gmd-12-4999-2019.

Sailor DJ. 2011. A review of methods for estimating anthropogenic heat and moisture emissions in the urban environment. *International Journal of Climatology* 31: 189–199. https://doi.org/10.1002/joc.2106.

Sailor DJ, Vasireddy C. 2006. Correcting aggregate energy consumption data to account for variability in local weather. *Environmental Modelling & Software* 21: 733–738. https://doi.org/10.1016/j.envsoft.2005.08.001.

Sanchez C, Williams KD, Collins M. 2016. Improved stochastic physics schemes for global weather and climate models. *Quarterly Journal of the Royal Meteorological Society* 142: 147–159. https://doi.org/10.1002/qj.2640.

Stewart ID. 2019. Why should urban heat Island researchers study history? *Urban Climate* 30: 100484. https://doi.org/10.1016/j.uclim.2019.100484.

Takane Y, Ohashi Y, Grimmond S, Hara M, Kikegawa Y. 2020. Asian megacity heat stress under future climate scenarios: impact of air-conditioning feedback. *Environmental Research Communications* 2: 015004. https://doi.org/10.1088/2515-7620/ab6933.

Tan J, Yang L, Grimmond CS, Shi J, Gu W, Chang Y, Hu P, Sun J, Ao X, Han Z. 2015. Urban Integrated Meteorological Observations: Practice and Experience in Shanghai, China. *Bulletin of the American Meteorological Society* 96: 85–102. https://doi.org/10.1175/BAMS-D-13-00216.1.

Titchner HA, Rayner NA. 2014. The met Office Hadley Centre Sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *Journal of Geophysical Research–Atmospheres* 119: 2864–2889. https://doi.org/10.1002/2013JD020316.

United Nation. 2019. *World Urbanization Prospects: The 2018 Revision*, New York: United Nations, Department of Economic and Social Affairs, https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf. Accessed 13th March 2020.

Vannière B, Demory M, Vidale PL, Schiemann R, Robert MJ, Co-authors. 2019. Multi-model evaluation of the sensitivity of the global energy budget and hydrological cycle to resolution. *Climate Dynamics* 52: 6817–6846. https://doi.org/10.1007/s00382-018-4547-y.

Vellinga M, Roberts M, Vidale PL, Mizielinski MS, Demory M-E, Schiemann R, Strachan J, Bain C. 2016. Sahel decadal rainfall variability and the role of model horizontal resolution. *Geophysical Research Letters* 43: 326–333. https://doi.org/10.1002/2015GL066690.

Vidale PL, Roberts MJ, Baker A, Bellouin N, Müller O, Schiemann R, Thornhill G, Vannière B. in prep. Description of the resolution hierarchy of HadGEM3-GC3.1 model as used in the HighResMIP atmosphere-only experiments. *Geoscientific Model Development*.

Walters D, Baran AJ, Boutle I, Brooks M, Earnshaw P et al. 2019. The Met Office unified model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geoscientific Model Development* 12: 1909–1963. https://doi.org/10.5194/gmd-12-1909-2019.

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

Wiltshire AJ, Duran Rojas MC, Edwards JM, Gedney N, Harper AB et al. 2020. JULES-GL7: the Global Land configuration of the Joint UK Land Environment Simulator version 7.0 and 7.2. *Geoscientific Model Development* 13: 483–505. https://doi.org/10.5194/gmd-13-483-2020.

Yin J, Yin ZE, Zhong HD, Xu SY, Hu XM, Wang J, Wu JP. 2011. Monitoring urban expansion and landuse/land cover changes of Shanghai metropolitan area during the transitional economy (1979–2009) in China. *Environmental Monitoring and Assessment* 177: 609–621. https://doi.org/10.1007/s10661-010-1660-8.

Zölch T, Maderspacher J, Wamsler C, Pauleit S. 2016. Using green infrastructure for urban climate-proofing: an evaluation of heat mitigation measures at the micro-scale. *Urban Forestry & Urban Greening* 20: 305–316. https://doi.org/10.1016/j.ufug.2016.09.011.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

APPENDIX A.

Offline JULES (vn. 5.4) simulations (Section 4.1) are conducted in a setup consistent with the (online) HadGEM3–PRIMAVERA climate simulations (GL7.0 science configuration; Wiltshire *et al.*, 2020). Two model configurations (CTRL–1T, HAD–1T) using the Best–1T urban model (Section 2.1; Best, 2005) are tested at the King's College London Strand Campus (KCL) site in central London (51.511°N, 0.117°W), for which hourly observations (forcing and evaluation data; Ward *et al.*, 2016) are available for 3 years (01/2011–12/2013). Instruments mounted on flux towers (49.6 m above ground level) are used to obtain the observations. Details about site characteristics and measurements are given in Kotthaus and Grimmond (2014a, 2014b). For an overview of JULES forcing/evaluation variables see Table 5 in Hertwig *et al.* (2020).

Table A1 gives essential parameters, data sources and settings for the two model configurations analysed. Differences between the settings focus on the land-cover fractions, bulk urban albedo (α), momentum roughness length (z_0) and anthropogenic heat emissions (Q_F). CTRL–1T uses land cover and morphology (z_0) derived from high-resolution (~1 m) GIS data for a 500-m radius source area around the site and an observed α. Q_F in CTRL–1T is supplied as monthly medians derived from hourly data of Kotthaus and Grimmond (2014a) generated with the GreaterQF model (Iamarino *et al.*, 2012) for central London (see Fig. 2 in Hertwig *et al.*, 2020). The HAD–1T run uses the HadGEM3–PRIMAVERA land cover of the grid-box containing the KCL site (note this is adjacent to the grid-box containing LWC and SJP; Table 1a) and GL7.0 default values for α, z_0 and Q_F. This setup differs from the Hertwig *et al.* (2020) study at the same site. Hertwig *et al.* (2020) used a different JULES science configuration, source area dependent land cover and the Cox (2001) phenology model for the Best–1T control run.

The model is spun up over the first year of the simulation until soil temperature and moisture have converged. Model output variables are hourly snapshots (at the model time step) and analysed as grid-box averages, that

TABLE A1 Parameters used for JULES offline simulations in central London (KCL site) for CTRL–1T and HAD–1T: (a) land-cover fractions, (b) urban parameters and (c) plant phenology. For further details see text and footnotes

	CTRL–1T	HAD–1T
(a) Land-cover fractions (%)		
Σ Vegetated	13[a]	0[b]
Broadleaf Tree	10	0
C3 Grass	3	0
Σ Non-vegetated	87[a]	100[b]
Urban (impervious)	66	98
Inland water	21	2
(b) Urban parameters		
Urban model	Best–1T (Best, 2005)	
Albedo (α)	0.11[c]	0.18[d]
Emissivity[d] (ε)	0.97	0.97
Heat capacity[d] (C)	0.28 MJ K^{-1} m^{-2}	0.28 MJ K^{-1} m^{-2}
Momentum roughness length (z_0)	1.76 m[e]	1.0 m[d]
Roughness ratio[d] (z_h/z_0)	10^{-7}	10^{-7}
Anthropogenic heat emissions (Q_F)	>0 W m^{-2}, monthly[f]	0 W m^{-2}, all[d]
(c) Phenology		
Leaf area index[d]	Monthly spatial fields	
Canopy height[d]	Monthly spatial fields	

[a]Derived from high-resolution (~1 m) London land-cover data.
[b]HadGEM3–PRIMAVERA model ancillary fields.
[c]Derived from observations (Kotthaus and Grimmond, 2014a).
[d]JULES–GL7.0 default value (Wiltshire *et al.*, 2020).
[e]Derived from London morphology (~1 m) data using the Macdonald *et al.* (1998) morphometric method.
[f]Derived from hourly GreaterQF model data for central London (Kotthaus and Grimmond, 2014a).

is, results from the JULES tiles in the grid-box are weighted by their plan-area fractions, as in the (online) HadGEM3–PRIMAVERA output. For the evaluation (Figures 3, S4), the model output is restricted to periods of available observations.