

Utility of thermal remote sensing for evaluation of a high-resolution weather model in a city

Thomas W. Hall^{1,2} | Lewis Blunn^{1,3} | Sue Grimmond¹ |
 Niall McCarroll¹ | Christopher J. Merchant^{1,4} | William Morrison^{1,5} |
 Jonathan K. P. Shonk³ | Humphrey Lean³ | Martin Best⁶

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

²Space Science Algorithmics (SPASCIA), Ramonville Saint-Agne, France

³Met Office@Reading, Reading, UK

⁴National Centre for Earth Observation, University of Reading, Reading, UK

⁵University of Freiburg, Freiburg, Germany

⁶Met Office, Exeter, UK

Correspondence

Thomas W. Hall, Space Science Algorithmics (SPASCIA), Ramonville Saint-Agne, France.
 Email: tom.hall@spascia.fr

Funding information

Met Office Strategic Priorities Fund Climate Resilience Programme -
 100 m-scale Modelling for Urban Climate Services: Urban Climate Services Using 100 m Resolution Modelling, Grant/Award Number: IUS-UM100; ERC Synergy Grant urbisphere, Grant/Award Number: 855005; NERC EOCIS, NE/X019071/1; NERC ASSURE, NE/W002965/1

Abstract

Progress in high-resolution numerical weather prediction (NWP) for urban areas will require new modelling approaches and extensive evaluation. Here, we exploit land surface temperature (LST) data from Landsat-8 to assess 100 m resolution NWP for London (UK) on four cloud-free days. The LST observations are directional radiometric temperatures with non-negligible uncertainties. We consider the challenges of informative comparison between the Landsat LST and the NWP scheme's internal characterisation of the complete surface temperature. The LST spatial coverage allows large-scale observation–model differences to be explored. In one case, obvious spatial artifacts in the NWP surface temperature are observed relative to the Landsat LST. These are found to be related to the NWP's initial method of downscaling of soil moisture using soil properties. Updated model runs have higher spatial correlation between model and Landsat LST. In cases where meteorological conditions favour the formation of horizontal convective rolls, warmer air temperatures associated with updraughts in the mixed layer extend inappropriately to the urban surface. This manifests as warm stripes in the model surface temperature that are not present in the Landsat LST. NWP–Landsat LST differences are larger in more built-up areas on days nearer summer solstice. This is largely attributed to urban thermal anisotropy, as Landsat preferentially views warmer urban surfaces, whereas the model LST represents all surfaces. We evaluate two approaches to quantify this sampling effect, but further work is needed to fully constrain it and facilitate more informative model evaluation.

KEY WORDS

land surface temperature, numerical weather prediction, thermal remote sensing, urban

1 | INTRODUCTION

High-resolution weather and climate modelling is of interest for day-to-day operations of cities and planning for future city conditions (Baklanov et al., 2018). Developing operational products for urban applications and services requires the development and evaluation of next-generation numerical weather prediction (NWP) models, with grid-cell resolutions of the order of 100 m being explored (e.g., Boutle et al., 2016; Lean et al., 2019). These scales pose new challenges as greater heterogeneity and complexity of urban form and properties are resolved, but they offer the potential to provide neighbourhood-scale information in support of a wide range of integrated urban services (World Meteorological Organization [WMO], 2019). To deliver the higher resolutions required in urban areas, sub-kilometre-scale models are being developed (Joe et al., 2018) to nest within kilometre-scale models.

The verification of models in urban areas remains a challenge given the lack of routine observations located appropriately (Grimmond & Ward, 2021). Any WMO observation sites (WMO, 2018a) in urban areas are likely to be located within the urban canopy layer, rather than the inertial sublayer or constant flux layer (Tang et al., 2021). Standard WMO site observations, such as those located in urban parks, represent grassed areas rather than the mix of buildings and vegetation that occur in different neighbourhoods. Citizen science weather stations—for example, Netatmo (Chapman et al., 2017; Fenner et al., 2021) and WOW (Kirk et al., 2021)—and WMO (2018b) recommended urban sites may better represent the mix of land covers in their source areas (Coney et al., 2022; Cornes et al., 2019; Muller et al., 2015; Vesala et al., 2008). If urban canopy layer observations are used for model evaluation, appropriate downscaling of variables from the inertial sublayer to within the urban canopy layer is required (e.g., Blunn et al., 2022; Tang et al., 2021; Theeuwes et al., 2019; Wang, 2014).

Eddy covariance (e.g., Hertwig et al., 2020; Masson et al., 2002) and large-aperture scintillometry (Saunders et al., 2024) sensors allow measurements of fluxes in the inertial sublayer, but with restricted horizontal spatial coverage (Grimmond & Ward, 2021; Ward et al., 2014) relative to a city. Ground-based remote-sensing techniques, such as automatic lidars and ceilometers and Doppler wind lidars, allow vertical profiles to be evaluated, but they remain limited in their horizontal coverage. Dense deployment of such sensors is generally limited to campaigns lasting months to years—for example, in the urbisphere in Berlin (Fenner et al., 2022) or in PANAME in Paris (Kotthaus et al., 2023).

Infrared imagers on satellites provide routine wide-area coverage during clear-sky conditions. There is a trade-off between return period and spatial resolution. A single Landsat satellite (with thermal resolution ~100 m) returns every 16 days at best (Malakar et al., 2018; Parastatidis et al., 2017), but more frequent ~1 km resolution observations have been used in urban areas (e.g., order 1 km; Leroyer et al., 2011).

Our objective here is to assess the use of high-resolution (100 m scale) thermal remote sensing from Landsat in evaluation of hectometre-scale NWP in urban areas. In the study we use the Met Office research model with a 100 m resolution (Lean et al., 2019), hereafter UM100 (Section 2). The Met Office routinely runs the so-called London Model, with an order 300 m resolution, nested within the UK variable-resolution (UKV) Unified Model (UM; Boutle et al., 2016). Here, UM100 runs are nested within a 300 m model (similar to the London Model) to explore and to help develop next-generation modelling on hectometre scales. The UM100 requires land cover plan-area fraction from which to determine the urban form and model parameters (e.g., albedo, emissivity). Given the crucial role of land cover information for both modelling and remote sensing, we also use a reference dataset to characterise the study area (Section 2.2). Consideration is given to the retrieval of land surface temperature (LST) from Landsat-8 (Section 3.2) and the implications of thermal anisotropy in urban areas (Section 3.3). An initial analysis of the model–observation differences on one day reveals some model improvements that could be addressed (Section 4.1), enabling revised UM100 runs for further, more detailed evaluation (Sections 4.1–4.5). This establishes the basic utility of thermal remote sensing in the evaluation of surface energy balance behaviour of hectometre-scale urban NWP models. Priorities for further work on methodologies for model–satellite comparisons are given in the concluding comments (Section 5).

2 | NWP METHODS AND STUDY AREA

2.1 | UM100 model runs

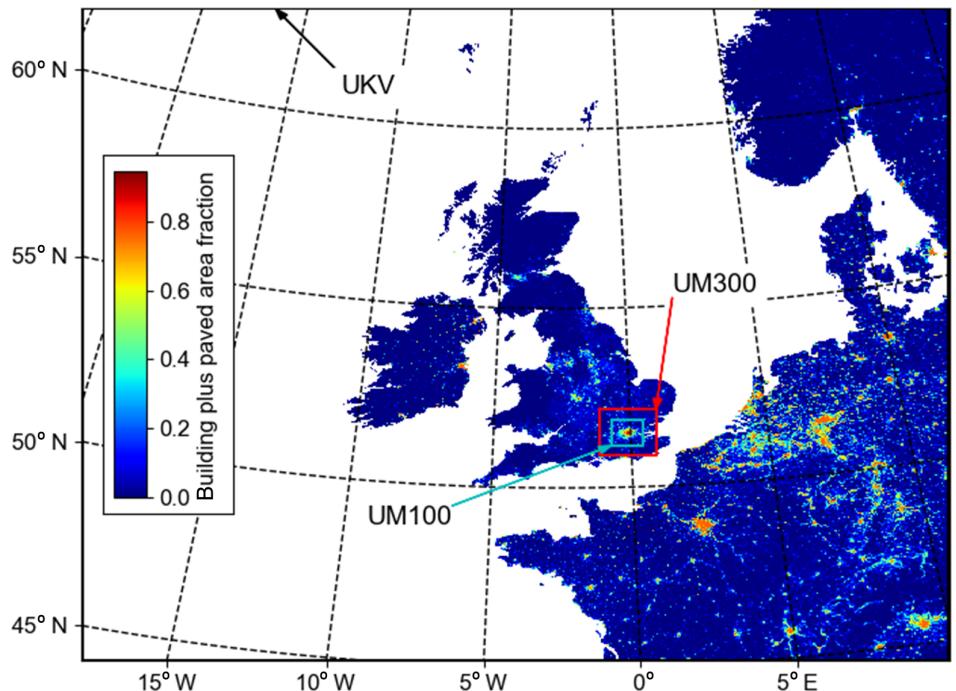
The UK Met Office's non-hydrostatic UM has a semi-implicit semi-Lagrangian numerical scheme for the deep-atmosphere dynamics (Davies et al., 2005; Wood et al., 2014). In this study, UM version 12.0 is run in research mode with one-way nesting to the highest resolution of 100 m (Lean et al., 2019), hereafter UM100 (Table 1). Four clear-sky days (Supporting Information Table S1) are analysed using 30 hr model runs, with the first 6 hr from 1800 UTC treated as spin-up (Table 1).

TABLE 1 Model grid resolutions and nesting. Figure 1 shows the areal extent of the three inner model domains.

Model	Horizontal grid length	Number of grid points	Vertical levels	Time step	Lateral boundary updated	Model start time (UTC)	Run length (hr)
Global	17 km	$1,536 \times 1,152$	70	12 min	n/a	1800	30
UKV	1.5 km	$950 \times 1,025$	70	1 min	60 min	1800	30
UM300	300 m	430×430	140	12 s	15 min	1800	30
UM100	100 m	800×800	140	3 s	15 min	1800	30

Note: Model start times are on the day before the case studies, allowing 6 hr spin-up.

FIGURE 1 Model domains extent for the UKV (outer map extent), UM300 (red), and the UM100 (blue) with building plus paved plan area fraction (colour bar) for the United Kingdom (Fuller et al., 1994) and Europe (Hollmann et al., 2013). Table 1 provides more details.



The Global Atmosphere and Land configuration 6.1 (Walters et al., 2017) is used for the outer global model (midlatitudes resolution 17 km; Table 1) to provide the boundary conditions for the UKV model (1.5 km resolution over the United Kingdom) (Tang et al., 2021). The UKV in turn provides the boundary conditions for the 300 m resolution model (UM300) (Boutle et al., 2016), from which the UM100 obtains its boundary conditions (Figure 1). The regional models use Regional Atmosphere and Land 3 (RAL3), which is the regional dynamics and physics scheme science configuration. Relative to the RAL2 baseline (Bush et al., 2022), the most significant evolution in RAL3 is the unification of the tropical and midlatitude configurations, with associated cloud and microphysics parametrisation updates (Bush et al., 2024; Field et al., 2023; Van Weverberg et al., 2021a, 2021b).

The UM100 80×80 km² domain has a 10–20 km buffer around Greater London (Section 2.2) to give the turbulent flow time to adjust to the higher resolution model grid (i.e., spin-up; Blunn, 2021; Lean et al., 2019). The subgrid

turbulence is parametrised using a “scale aware” blending scheme (Boutle et al., 2014). The local three-dimensional Smagorinsky–Lilly turbulence scheme (Halliwell, 2007; Lilly, 1962; Smagorinsky, 1963) and the non-local one-dimensional planetary boundary layer (BL) scheme (Lock et al., 2000) are weighted, such that the local scheme is preferred with decreasing horizontal grid length.

The Joint UK Land Environment Simulator (JULES) version 6.1 land surface model (Best et al., 2011; Clark et al., 2011) urban option MORUSES (Bohnenstengel et al., 2011; Porson et al., 2010a, 2010b) is used. MORUSES has improved timing and amplitude of the sensible heat flux in central London (Bohnenstengel & Hendry, 2016; Hertwig et al., 2020; Warren et al., 2018), with respect to the Best (2005) urban option used by Lean et al. (2019). Anthropogenic heat flux is assumed to vary monthly but with no diurnal or daily variations (Hertwig et al., 2020). The values are derived from monthly mean energy consumption for 1995–2003 (DUKES, 2003). The flux is assumed to vary with built (or impervious) fraction

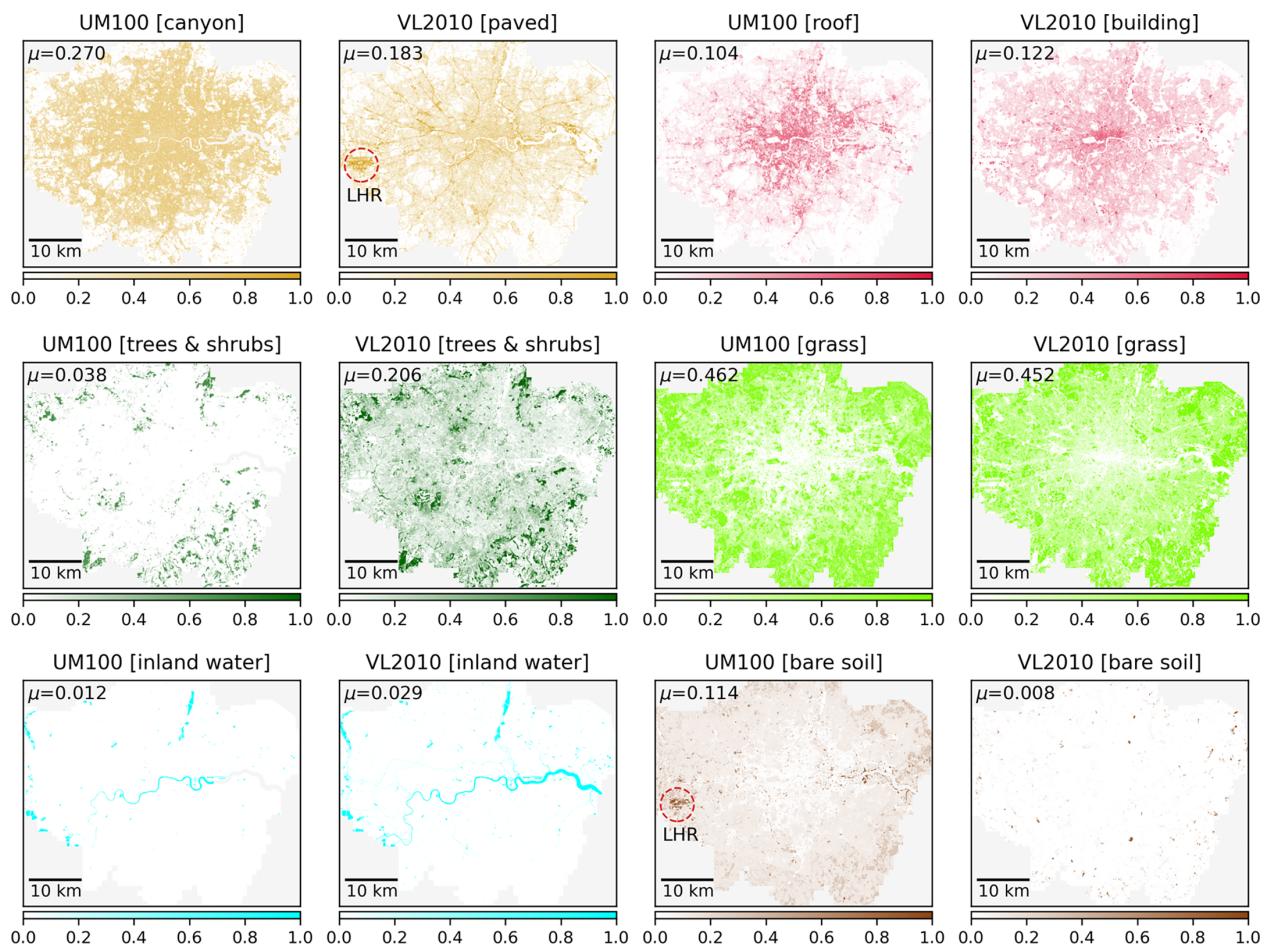


FIGURE 2 Greater London (Figure 1 shows location in the United Kingdom) land cover fractions at 100 m resolution from two sources (Table 2), with higher fractions having darker shading. The mean fraction μ is given in each case. Note UM100 bare soil fractions for London Heathrow Airport (LHR, circled) compared with VL2010 paved.

between 16.7 and 26.2 W·m⁻² (for 100% impervious grid cell) for different months of the year.

2.2 | Study area

Greater London (Figure 2), covering 1,569 km², experiences a humid warm temperate climate (Cfb) (Kottek et al., 2006). To characterise the UM100 model domain (Figure 1), each grid cell is assigned the relevant land cover fraction for the different JULES tile types (Table 2). Operationally, the Met Office uses the so-called Institute of Terrestrial Ecology (ITE; now Centre for Ecology and Hydrology) surface classes derived from Landsat-5 Thematic Mapper data (Fuller et al., 1994). For MORUSES (Section 2.1), the “urban” ITE class is split into canyon and roof using empirical relations (Bohnenstengel et al., 2011), and other JULES tiles (vegetated, bare soil). A number of issues arise from use of the ITE dataset: it is outdated (Bunce et al., 1990) relative to the continually changing

urban landscape; it has coarse resolution (25 m) relative to the sizes of buildings, roads, and trees; and the Bohnenstengel et al. (2011) relations are applied at a very different resolution (100 m) from their development (1 km).

To explore this further, the UM100 surface cover is compared with a more recent, higher resolution dataset, hereafter referred to as VL2010 (Lindberg & Grimmond, 2011a, 2011b; Lindberg et al., 2018). Based on 1 m observations aggregated to 4 m, VL2010 has five land cover classes. Here, all trees and shrubs are collapsed into one class (Table 2). To compare the datasets (UM100, VL2010) the projections are first made consistent. This involves reprojecting the ITE data onto the rotated pole grid of the UM and making the necessary datum correction.

Comparison of the land cover datasets (Figure 2) indicates the UM100 canyon fraction is overestimated and the roof fraction is underestimated, particularly outside the city centre. The UM100 trees-plus-shrubs fraction is underestimated for the Greater London area. Bare soil

TABLE 2 Surface cover fractions are used to weight the JULES calculated variables for each tile within a UM100 grid cell. The surface cover is compared with a reference dataset (VL2010)—Lindberg and Grimmond (2011a) and Lindberg et al. (2018) provide details—which combines classes as indicated.

JULES tile-UM100	VL2010
Broadleaf trees	Trees and shrubs
Needleleaf trees	
Shrubs	
C3 grass	Grass
C4 grass	
Inland water	Water
Bare soil	Bare soil
Canyon	Paved surface
Roof	Buildings or roofs

is generally overestimated, and is particularly unrealistic over Heathrow Airport (LHR, circled in Figure 2). As these land cover fractions are the basis for assigning model parameters (Bohnenstengel et al., 2011; Hertwig et al., 2021), they impact the UM100 modelled surface temperature (Section 2.3).

2.3 | UM100 surface temperature

The UM100 surface temperatures are derived from the JULES surface energy balance fluxes (Best et al., 2011) for each tile (Table 2) with the canyon and roof tiles using the MORUSES option (Best et al., 2011; Porson et al., 2010a). Tile emissivities $\epsilon_{s,t}$ have temporally and spatially fixed values except for the canyon tile, which varies with canyon geometry using fixed material values for wall and road. The grid-cell surface emissivity ϵ_s is a function of the tile surface fractions f_t :

$$\epsilon_s = \sum_t f_t \epsilon_{s,t}. \quad (1)$$

The surface emissivity in the UM100 is compared with the 100 m resolution Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Emissivity Dataset (ASTER GEDv3; Hulley et al., 2015) used in the Landsat LST retrieval (Section 3.2) on the UM grid (Supporting Information Figure S2). The mean UM100 emissivity (0.964) is lower than ASTER GED (mean of 0.972). The Landsat LST sensitivity to this difference is $\sim 0.4\text{--}0.6$ K for the four days studied. Unrealistically low emissivity values (~ 0.90) for LHR come from the high bare-soil fractions assigned (Figure 2).

The grid-cell value surface temperature T_S is a function of the tile surface fractions and the tile surface temperatures $T_{s,t}$:

$$T_S = \sum_t f_t T_{s,t}. \quad (2)$$

The hourly model temperatures are linearly interpolated in time to the Landsat overpass time (~ 1100 UTC, Supporting Information Table S1) to give “UM100 LST”. The Landsat and UM100 LST are compared within the VL2010 domain (Figure 2) with a common mask. The UM100 minus Landsat LST difference is designated Δ LST. In Figures 5 and 6 (Section 4.1), maps of Δ LST are presented with the mean Δ LST subtracted (zero-mean differences) to highlight spatial patterns of variability.

3 | LANDSAT METHODS

3.1 | Landsat-8 observations

Landsat-8, launched in 2013, is in a sun-synchronous, near-polar orbit with a 16-day repeat cycle (USGS, 2021a). The London overpass time is around 1100 UTC for the dates relevant to this study (Supporting Information Table S1). Landsat-8 carries two sensors: the Operational Land Imager with nine spectral bands in the visible to shortwave infrared wavelengths; and a thermal infrared sensor (TIRS) with two spectral bands in the atmospheric window ($10.6\text{--}11.2\mu\text{m}$ and $11.5\text{--}12.5\mu\text{m}$) (USGS, 2021a).

The Landsat Collection 2 Level-1 data products (USGS, 2021b) including the operational land imager data (nominal resolution 30 m) are used to assess cloud cover and to estimate surface albedo. To identify candidate clear-sky days, the Landsat total cloud cover fraction cloud mask derived from the quality assurance band (USGS, 2021b) is initially used, followed by assessment of the morning cloud optical depth estimated from hourly UKV profiles of liquid and ice water mass fractions. Four study dates are selected (Supporting Information Table S1) because of their near-zero ($<4\%$) total cloud cover in the Landsat image and very low total cloud optical depths (<0.05) from 0900 UTC until the observation time across the UM100 domain.

The Landsat data are transformed from the WGS84 coordinate reference system to the British National Grid, then reprojected and interpolated to the rotated pole UM100 grid. Supporting Information Figure S3 shows a false-colour visible image, cloud mask, and band 10 (TIRS) brightness temperature on one of the study dates (July 15, 2018) after transforming the Landsat data to the UM100 grid.

TABLE 3 Comparison of three retrieval methods using Landsat-8 thermal infrared sensor (TIRS) data to obtain land surface temperature.

	This study	FORTH (Parastatidis et al., 2017)	NASA JPL (Cook et al., 2014)
Atmospheric profiles	ERA5 ($0.25^\circ \times 0.25^\circ$) (Hersbach et al., 2020)	NCEP–NCAR ($2.5^\circ \times 2.5^\circ$) (Kalnay et al., 1996)	GEOS-5 FP-IT ($0.625^\circ \times 0.5^\circ$) (Lucchesi, 2013)
RTM	RTTOV-13 (Saunders et al., 2020)	MODTRAN4 (Berk et al., 1999)	MODTRAN4 (Berk et al., 1999)
Algorithm	Equation (3)	Jimenez-Munoz et al. (2009, 2014)	RIT–JPL (Cook et al., 2014)

Abbreviations: ERA5, European Centre for Medium-Range Weather Forecasts Reanalysis v5; FORTH, Foundation for Research and Technology Hellas; GEOS-5 FP-IT, Goddard Earth Observing System Model Version 5, Forward Processing for Instrument Teams; JPL, Jet Propulsion Laboratory; NASA, National Aeronautics and Space Administration; NCEP–NCAR, National Centers for Environmental Prediction–National Center for Atmospheric Research; RIT, Rochester Institute of Technology; RTM, radiative transfer model.

3.2 | LST retrieval and assessment

Here, we apply a single-channel radiative-transfer-based algorithm to retrieve LST from Landsat-8 TIRS data (Table 3, Figure 3). Band 10 ($10.6\text{--}11.2\,\mu\text{m}$) is used as it has the lower calibration uncertainty of the two thermal bands (USGS, 2019a). Its nominal resolution is 100 m (USGS, 2021a), and a “real” resolution of around 115 m has been estimated (USGS, 2019b).

The radiative transfer model (RTM) RTTOV (Saunders et al., 2020) is used to calculate the downwelling and upwelling radiance and the atmospheric transmittance for calculating the blackbody radiance from the land surface $B_\lambda(T_{\text{LST}})$ using

$$B_\lambda(T_{\text{LST}}) = \frac{L_{\text{TOA}}(\lambda) - \sum_i B_\lambda(T_i)\Delta t - (1 - \varepsilon)t_s \sum_i \frac{B_\lambda(T_i)\Delta t}{t_{i-0.5}t_{i+0.5}}}{\varepsilon t_s}, \quad (3)$$

where L_{TOA} is the top-of-atmosphere radiance observed by Landsat, $B_\lambda(T_i)$ is the blackbody radiance from level i , t_i is the level i to space transmittance, t_s is the surface to space transmittance, and ε is the surface emissivity obtained from ASTER GEDv3 (Hulley et al., 2015). The mean of ASTER band 13 ($10.25\text{--}10.95\,\mu\text{m}$) and band 14 ($10.95\text{--}11.65\,\mu\text{m}$) emissivity is used as the Landsat band 10 emissivity. Atmospheric profiles of temperature, water vapour, carbon dioxide, and ozone from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5; Hersbach et al., 2020) interpolated to the Landsat observation times and locations are provided as input to RTTOV (Table 3). The land surface blackbody radiance is converted to LST using a look-up table with 0.01 K precision, used to obtain the LST by interpolating between the radiance values closest to $B_\lambda(T_{\text{LST}})$.

To verify the retrieved LSTs and gauge their uncertainty, we compare the retrieved LST with two Landsat LST products based on different algorithms (Table 3) that also use ASTER GEDv3 emissivities. MODTRAN4 (Berk

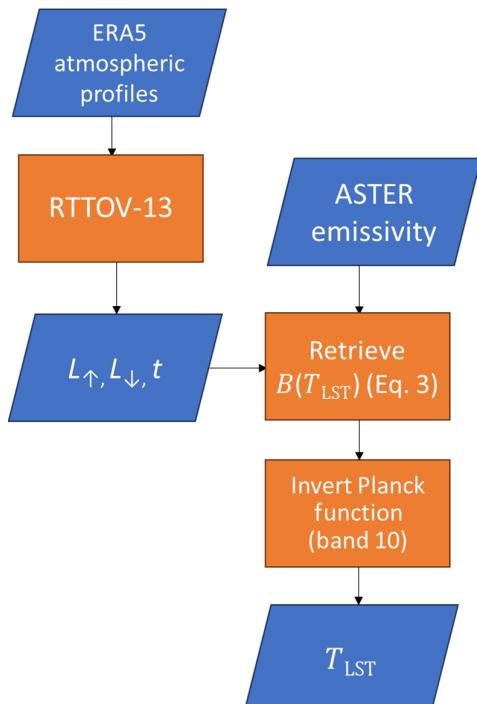


FIGURE 3 Summary of Landsat land surface temperature (T_{LST}) retrieval algorithm with inputs/outputs (blue parallelograms) and processes (orange rectangles) indicated, where L_\uparrow and L_\downarrow respectively refer to the upwelling and downwelling long-wave radiance and t is atmospheric transmittance. Abbreviations: ASTER, Advanced Spaceborne Thermal Emission and Reflection Radiometer; ERA5, European Centre for Medium-Range Weather Forecasts Reanalysis v5.

et al., 1999) is used in both the Foundation for Research and Technology Hellas (FORTH) and Jet Propulsion Laboratory (JPL) algorithms but with different inputs (Table 3). FORTH uses over 4,000 simulations with different atmospheric profile inputs (Jimenez-Munoz et al., 2009, 2014), allowing single-equation operational global-scale LST. JPL (Cook et al., 2014) uses Goddard

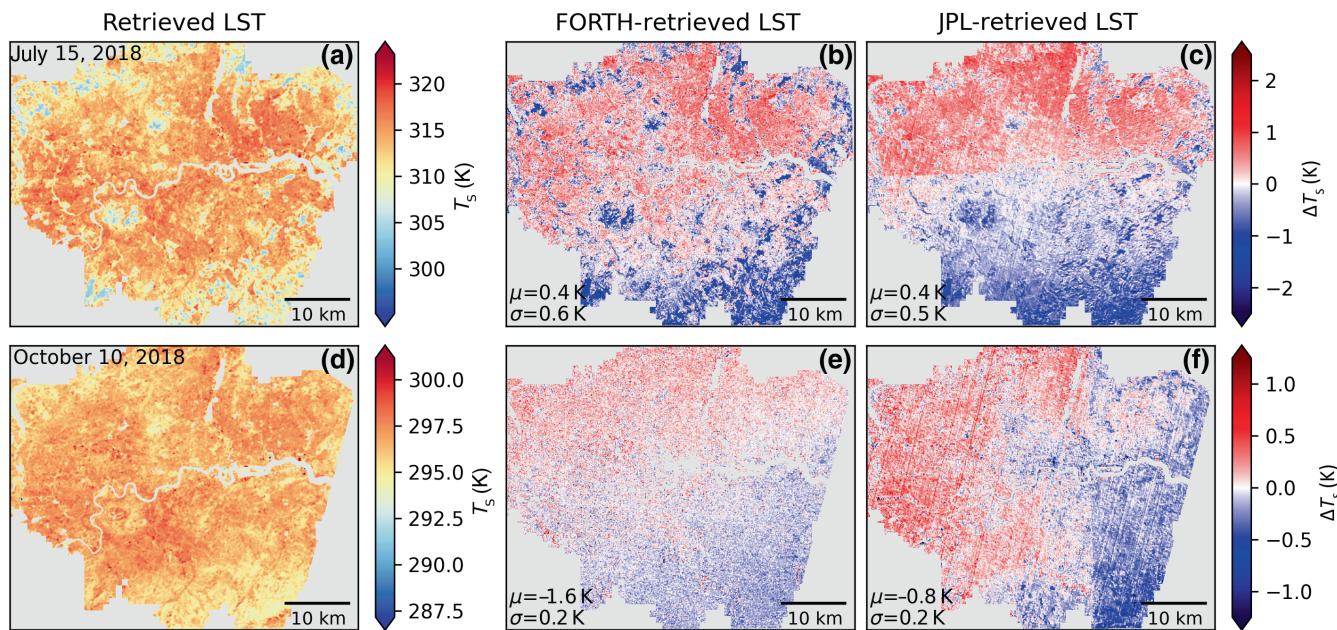


FIGURE 4 Land surface temperature (LST) products (Table 3) for (a–c) July 15, 2018, and (d–f) October 10, 2018, retrieved using (a, d) this study and the differences (b, e) Foundation for Research and Technology Hellas (FORTH) minus this study and (c, f) Jet Propulsion Laboratory (JPL) minus this study. Panels (b), (c), (e), and (f) are zero meaned; that is, the mean difference μ is subtracted before plotting. The mean difference μ and standard deviation σ values are given in each case. Statistics are calculated after applying a common spatial mask to each dataset.

Earth Observing System Model Version 5, Forward Processing for Instrument Teams (GEOS-5 FP-IT) reanalysis data (Lucchesi, 2013) and L2 top-of-atmosphere reflectance data.

The mean difference μ is removed to highlight spatial patterns of variability of the three retrievals (Figure 4). The size of the mean LST differences (-1.6 to $+0.4$ K) and their standard deviations (σ , 0.2–0.6 K) are of the order expected given the radiometric uncertainty from Landsat-8 (Gerace & Montanaro, 2017). Differences are indicative of the uncertainty arising from the different retrieval methods, with the median uncertainty in LST from the surface emissivity uncertainty being ~ 0.3 K. Other uncertainty arises from different atmospheric profile data used (Table 3). Visible discontinuities in the JPL-retrieved LST differences occur due to changes in the JPL retrieval input profiles at the grid-cell boundaries at 51.5°N (Figure 4c) and 0° longitude (Figure 4f). The total uncertainty (~ 1 K) is an order of magnitude less than the observed spatial and temporal variability. The Pearson correlation coefficient calculated between the three datasets is above 0.97 on both days (Supporting Information Table S2), again indicating uncertainty in the observed LST is much smaller than the spatial variability.

From comparison of the three Landsat LST retrievals, we conclude that model-observation differences

(Section 4.1) greater than the observed ~ 1 K differences between retrievals (Figure 4) can be attributed to the NWP model (e.g., physics, ancillaries) and sampling effects (thermal anisotropy, Section 3.3), rather than observation errors.

3.3 | Comparison of Landsat with UM100 LST

The viewing zenith angle of Landsat varies from 0° (nadir) to $\pm 8.6^{\circ}$. This near-nadir view results in near-horizontal facets being relatively heavily weighted in the sensor field of view. As vertical facets (i.e., walls) make up a large fraction of the total surface area in cities, the “complete” surface temperature should account for them in both observations and models (Voogt, 2000; Wang et al., 2022). However, neither satellite views nor urban land surface models do this, and they differ in their facet weightings and definitions. In late morning under clear skies, near-horizontal buildings’ facets (e.g., roofs) can be significantly warmer than walls that have been shaded within urban canyons (Morrison et al., 2021). This qualitative understanding of the sampling or definitional difference due to thermal anisotropy predicts that Landsat will be warmer than the model LST on sunny days in more urbanised areas.

There are two logical approaches to minimise this definitional difference: extract a nadir-view LST from the model or obtain a complete surface temperature T_c from Landsat. The former approach is challenging, as the canyon tile combines paved surfaces (e.g., roads) and walls requiring the horizontal (paved + roofs) and vertical (walls) facet temperatures to be separated before combining with appropriate vegetated surface values. Here, we consider the latter approach, estimating T_c from the Landsat LST T_r based on geometric and meteorological information (Yang et al., 2020) (Section 4.3).

We also compare LST retrieved from the moderate resolution imaging spectroradiometer (MODIS) on one day (July 15, 2018) also to the UM100 (Section 4.4). The MOD11A1 Version 6.1 Level-3 1 km resolution gridded LST product (Wan et al., 2021) is obtained using a generalised split-window LST algorithm (Wan & Dozier, 1996). As with the Landsat data, the MODIS LST data are transformed from their native (sinusoidal) projection to the British National Grid, then transformed to the UM100 grid. On this study day, MODIS observed London with an oblique view of $\sim 52^\circ$ at 1200 UTC, resulting in a greater weighting of vertical facets and therefore an LST more representative of T_c (Jiang et al., 2018). However, the MODIS spatial resolution (1 km) is much coarser than Landsat (~ 100 m).

4 | RESULTS AND DISCUSSION

Comparison of the UM100 and Landsat LST identified issues related to surface ancillary data (treatment of soil moisture, leaf area index, and spatial mismatch of surface mapping due to incorrect co-ordinate reference systems). Here, we present UM100 runs illustrating the soil moisture issue only. All runs shown here used RAL3 (Section 2.1) after leaf area index and spatial mismatch issues were

addressed, having been highlighted in earlier (RAL2) runs and comparisons. In the “initial” runs shown here, soil moisture issues remain, and in the “final” runs the soil moisture treatment is updated.

4.1 | Spatial assessment of UM100 LST

Initial model run results for July 15, 2018, show an overall model cold bias with mean difference in LST of 6.4 K (Figure 5). There is a clear lack of spatial agreement between the observed and modelled LST: the high UM100 LSTs above 315 K in the southeast of the domain do not have a corresponding analogue in the observations, and the standard deviation of the differences is high, at 3.4 K. The high UM100 LSTs are associated with a “blocky” LST pattern that also extends more subtly south and northeast of Greater London (Supporting Information Figure S4d) and correlates with low model soil moisture (Supporting Information Figure S4a). Throughout the UM100 domain, low and high soil moisture patterns correlate with high and low LST respectively (Supporting Information Figure S4). As these LST patterns are not found in Landsat, we conclude that the soil moisture patterns are unrealistic.

In the initial model run, UM100 soil moisture (Supporting Information Figure S4a) is calculated by downscaling the global model soil moisture stress using 100 m resolution soil ancillaries (critical point and wilting point). The 100 m critical point and wilting point fields contain sharp gradients, with regions of low critical point and wilting point correlating with regions of low UM100 soil moisture. To ameliorate the issue, final model runs use the global critical point and wilting point linearly interpolated to the 100 m grid, resulting in a smoother soil moisture field without regions containing low values (Supporting Information Figure S4b).

The final model runs (Figure 6) show that the southeast hotspot in July 2018 is largely removed (cf. Figure 5).

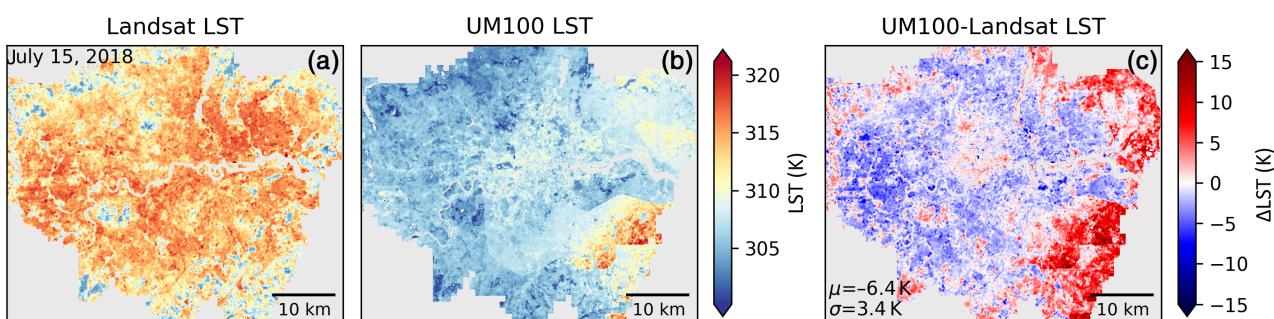


FIGURE 5 Land surface temperature (LST) on July 15, 2018, from (a) Landsat and (b) UM100 (initial run), and (c) (UM100 – Landsat) zero-mean (mean subtracted) difference, with mean difference μ and standard deviation σ calculated for the common area in each dataset.

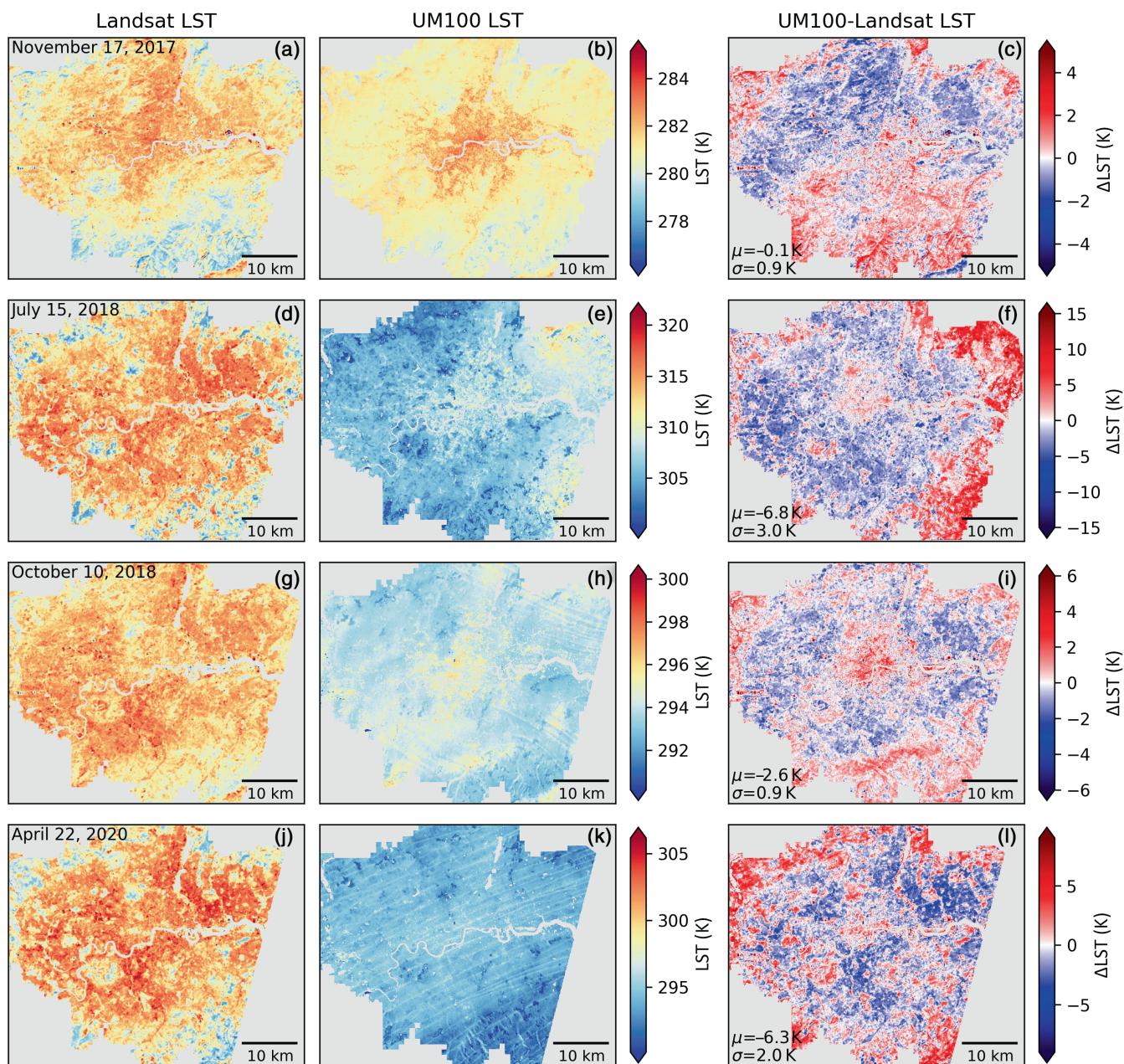


FIGURE 6 Land surface temperature (LST) on four days from (a, d, g, j) Landsat and (b, e, h, k) UM100 (final runs), and (c, f, i, l) their zero-mean (mean subtracted) difference, with mean difference μ and standard deviation σ calculated for the common area in each dataset.

This results in a spatial distribution of UM100 LST closer to the observations, with reduced standard deviations of the differences from 3.4 K to 3.0 K. However, the mean difference increases from -6.4 K to -6.8 K . This can be attributed to an average soil moisture increase in the new model runs (Supporting Information Figure S4c). The October ($\mu = -2.6 \text{ K}$) and November ($\mu = -0.1 \text{ K}$) cases have a relatively small overall model–Landsat bias, whereas a strong cold bias is present in the April ($\mu = -6.3 \text{ K}$) and July ($\mu = -6.8 \text{ K}$) cases. The exact cause of this seasonal variation is not explored here, but it

is partly associated with April and July being closer to the Northern Hemisphere summer solstice, when there is more short-wave radiation. In such circumstances, larger energy fluxes and temperature gradients develop between the land surface, atmosphere, and soil, with the potential for greater errors. The standard deviations of the differences are also much smaller in the October (0.9 K) and November (0.9 K) cases compared with the April (2.0 K) and July (3.0 K) cases. Although each day has different spatial patterns in the differences, larger positive values tend to occur in the more

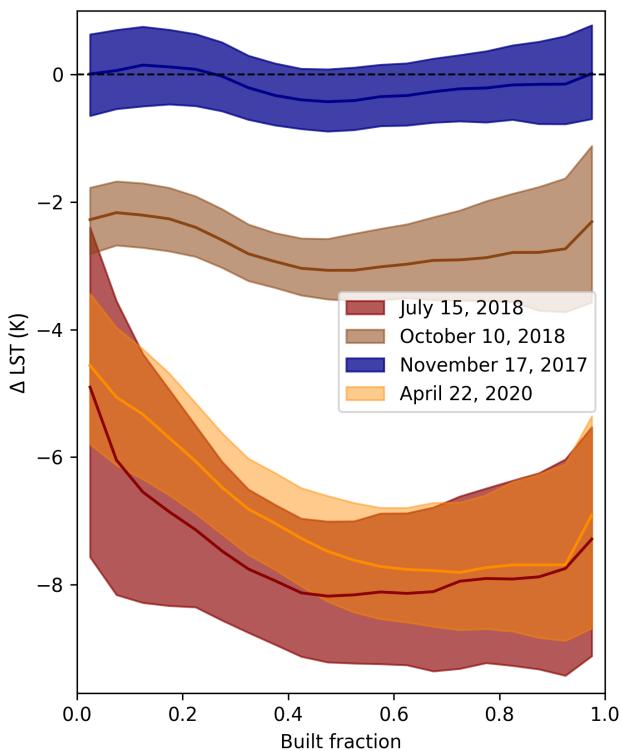


FIGURE 7 Variation of median (line) and interquartile range (shading) ΔLST (equal to UM100 minus Landsat) with built (paved + building) fraction on four days (key). Built fractions are from the VL2010 dataset (Table 2), with grid cells with a water fraction greater than 0.25 removed from this analysis.

vegetated areas, particularly on the outskirts of Greater London compared with the more densely built-up interior (Section 4.3). The diagonal stripes visible (Figure 6) in the UM100 LST but not Landsat in April 2020 are discussed in Section 4.5.

4.2 | Comparison to air temperature

As a general assessment of the final runs, we use 1.5 m air temperature observations (Met Office, 2012) from five sites (Heathrow Airport, Kew Gardens, St James Park, Kenley Airfield, Northolt) across Greater London. We find a model air temperature cold bias of -0.6 to -4.1 K across the four case study days. This is consistent in terms of sign with the model–Landsat LST cold bias; however, across dates and sites there is considerable scatter in these air–surface bias results (Supporting Information Figure S5). This reflects the different processes controlling the LST (cf. air temperature) bias, which likely includes critical impacts from model–Landsat LST definitional differences (Section 3.3). The relative model–Landsat LST cold bias is explored further in Sections 4.3 and 4.4.

4.3 | Relation of LST differences to land cover

4.3.1 | Built fraction

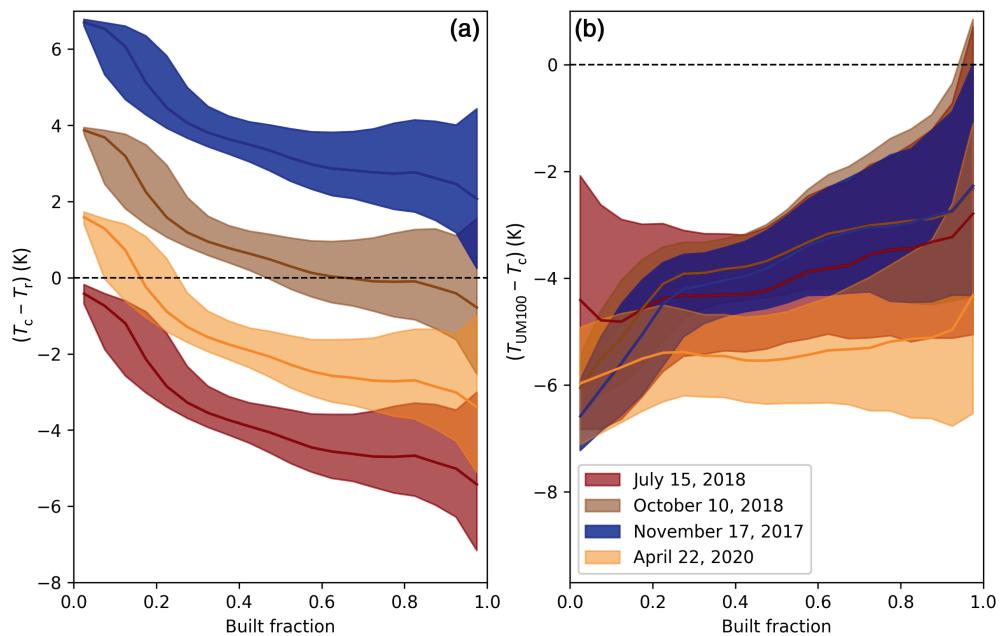
To better understand the controls on the spatial distribution of model–observation differences (ΔLST), differences are analysed against the built-area (paved + building) fractions from VL2010 (Figure 2). The dependencies are shown for the four days in Figure 7. The UM100 LST becomes increasingly cold relative to Landsat with greater built fraction (lower vegetated fraction) on all four days. This trend is much stronger on the two days (July 15, 2018 and April 22, 2020) with the larger mean ΔLST . This indicates that there is a larger temperature contrast between built and highly vegetated areas, such as parks, in the Landsat imagery relative to UM100 on these two days. This tendency for larger LST differences with increasing built fraction may be attributed to the sensor field of view or sampling effect of Landsat described in Section 3.3. To mitigate this effect, we estimate the complete surface temperature T_c from the Landsat LST T_r , adding to T_r the adjustment from Yang et al. (2020).

The Yang et al. (2020) adjustment comprises a model based on three-dimensional simulations of urban surface facet temperatures for different surface geometries and properties, meteorological conditions, and solar angles. In their study area (Hong Kong), daytime differences between T_c and T_r reach up to 10 K, varying with geometric characteristics such as plan area fraction of buildings λ_p (Supporting Information Figure S1) and wall facet area to building footprint area fraction F , and meteorological conditions (viz., solar irradiance, wind speed). Yang et al. (2020) report greater differences between T_c and T_r in more built-up areas (i.e., larger both λ_p and F) and with higher solar irradiance.

We calculate the Yang et al. (2020) adjustment ($T_c - T_r$) for London. The magnitude and sign of the adjustment varies between days (Figure 8a), because of differences in mean LST, solar irradiance, and solar geometry. The adjustment becomes cooler at higher built fractions. After applying this correction, the differences between ΔLST on each day (Figure 8b) are greatly reduced and the overall trend with built fraction for July 15, 2018, and April 22, 2020, reduces and changes sign. However, a model cold bias of 3–7 K (median) for each day remains.

Our use of the Yang et al. (2020) empirical method to estimate T_c from the Landsat LST T_r entails several caveats: impacts of vegetation on temperature are not accounted for in this model; their parameter ranges do not cover all our cases (e.g., low winter sun angles at London latitudes); the material types and canopy geometries considered in Hong Kong are likely to be different to London,

FIGURE 8 As Figure 7, but using the complete temperature T_c derived using the Yang et al. (2020) empirical adjustment added to the Landsat LST T_r . (a) The values of the adjustment. (b) UM100 LST (T_{UM100}) minus T_c (shown with the same vertical scale as Figure 7).



which will affect the relation between T_c and T_r . More work is needed to obtain locally applicable parameters for the London area.

4.3.2 | Dominant land cover

To explore the role of model land cover type we analyse ΔLST ($T_{UM100} - T_r$) for the grid cells with one “dominant” surface type with a fraction >0.5 (Figure 9). On July 15, 2018, and April 22, 2020, there are consistently larger ΔLST (Figure 7) and larger median ΔLST for all four VL2010 built and vegetated classes (Figure 9). There are small differences up to 0.4 K on all four days between the median ΔLST for the two built classes (vertical red and gold lines, Figure 9). This implies that the model–observation bias is relatively insensitive to whether a surface is predominantly roof or paved. The differences in median ΔLST are larger between the two vegetated classes between days (0.4–3.0 K), with the July 2018 case having a median ΔLST of -4.1 K for trees and shrubs but -7.1 to -7.7 K for grass, roof, and paved classes (Figure 9b). The median ΔLST is smallest for trees and shrubs on all days, except when the differences are very small (<0.2 K) for all classes on November 17, 2017 (Figure 9d). The areas with trees and shrubs as the dominant class also tend to have the lowest Landsat and model LST (Figure 6).

However, the definitional difference between the Landsat LST and the model analogue complicates the interpretation of these results. If we take $\Delta LST = T_{UM100} - T_c$ (Supporting Information Figure S6), the differences between the median ΔLST for the two built classes increase to 3.5–3.9 K. The median ΔLST becomes

much smaller for roofs on all days except November 17, 2017 (Supporting Information Figure S6c), with much smaller changes for the other three classes. This is because the areas where roofs dominate occur mainly in central London (Figure 9a), where larger differences between T_c and T_r are expected linked to their larger λ_p and F values.

4.4 | Assessment of UM100 LST with MODIS

Comparison of the MODIS LST with the updated model run on July 15, 2018 (Figure 10) at 1200 UTC has both a much smaller mean difference (-0.7 K cf. -6.8 K for UM100–Landsat) and smaller standard deviation (2.6 K cf. 3.0 K). This smaller mean difference is largely explained by MODIS’s oblique view angle of $\sim 52^\circ$ and the relation between the satellite and solar azimuth angles. Simulations of thermal anisotropy in central London on August 27, 2017 (Morrison et al., 2023), relative to a nadir view, show the highest LST is obtained when the satellite and solar azimuth angles align, whereas the lowest LST is obtained when these angles are 180° apart, as the satellite will view more shaded facets. This effect increases as the satellite view angle increases away from nadir. On July 15, 2018 (Figure 10), the MODIS view azimuth angle of -67° is 116° from the solar azimuth angle φ_s of $+177^\circ$. As this sun–sensor azimuth angle difference exceeds 90° , when combined with its oblique view angle, MODIS preferentially views shaded (cf. sunlit) facets, resulting in a significantly lower LST than Landsat’s near-nadir ($<5^\circ$) view.

Jiang et al. (2018) found, based on a single satellite view angle, the closest match to the complete urban surface

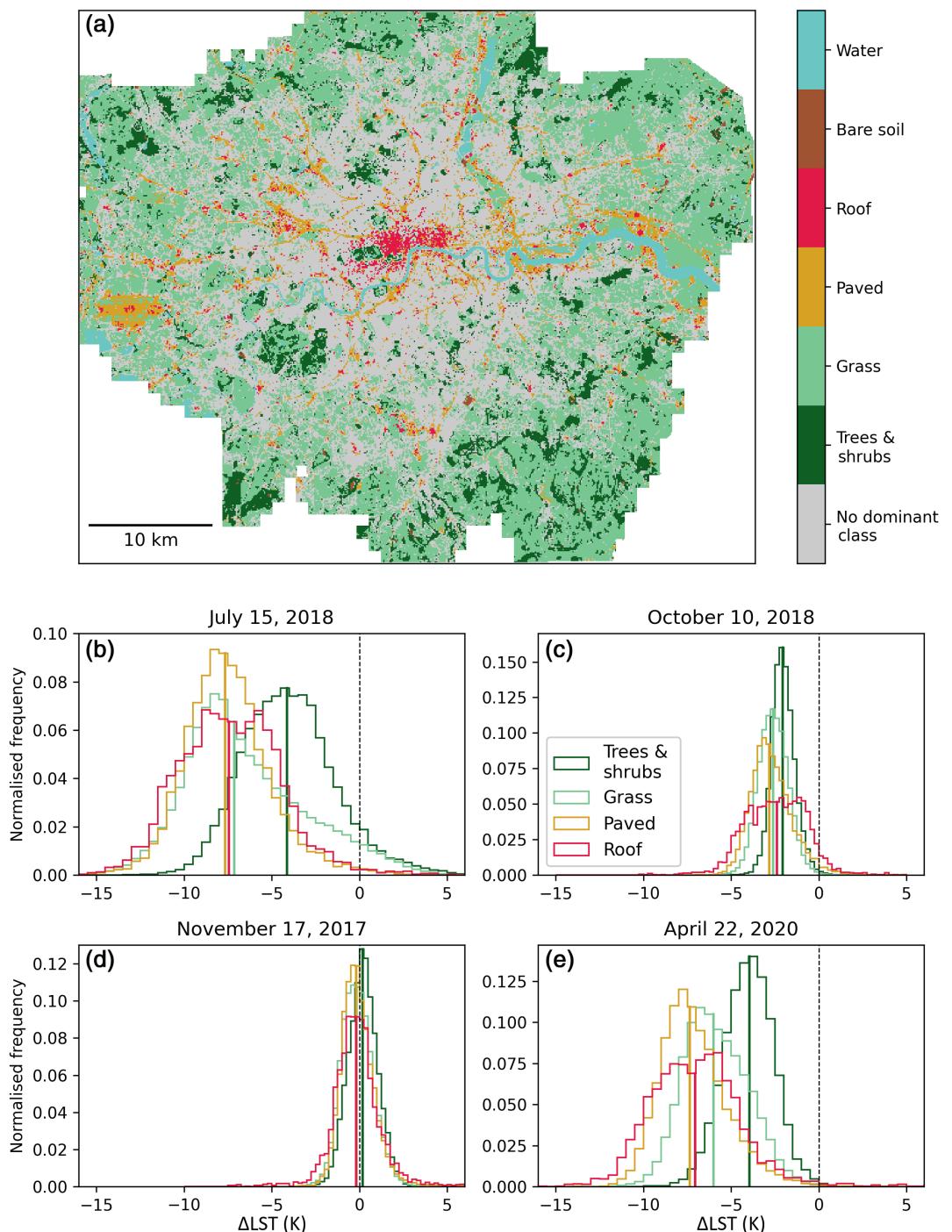


FIGURE 9 Analysis of ΔLST (UM100 minus Landsat) for UM100 grid cells with one dominant land cover type (>0.5 , based on VL2010): (a) spatial pattern of dominant class (colour bar); (b–e) normalised distribution of ΔLST by land cover type (key: bare soil not shown) with median (vertical lines), (b, e) 0.5 K and (c, d) 0.25 K bins for (b) July 15, 2018, (c) October 10, 2018, (d) November 17, 2017, and (e) April 22, 2020. See also Supporting Information Figure S6.

temperature is obtained with a satellite azimuth angle of $\varphi_s \pm 90^\circ$ and zenith angle of 45–60°. This suggests that the MODIS LST is more representative of the complete surface temperature than Landsat in this case. Given the MODIS–solar azimuth difference, a small cold bias relative to the complete surface temperature is expected instead of the warm bias for Landsat.

Despite the similar mean LST, there are clear differences in the spatial distribution between MODIS and UM100. The spatial distribution of the differences appears similar to the corresponding UM100–Landsat differences (Figure 6f). The relative model warm bias to the east of the domain is still present, and there is a noticeable ~5 K warm bias in central London.

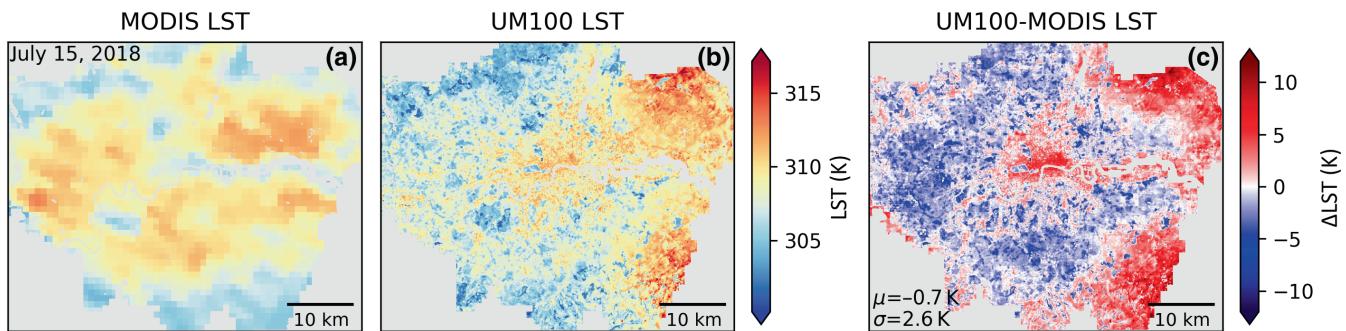


FIGURE 10 Comparison for July 15, 2018, at 1200 UTC of land surface temperature (LST) from (a) MODIS and (b) UM100, and (c) their zero-mean difference (UM100 minus MODIS), with the mean difference μ and standard deviation σ calculated for the common area of the two datasets.

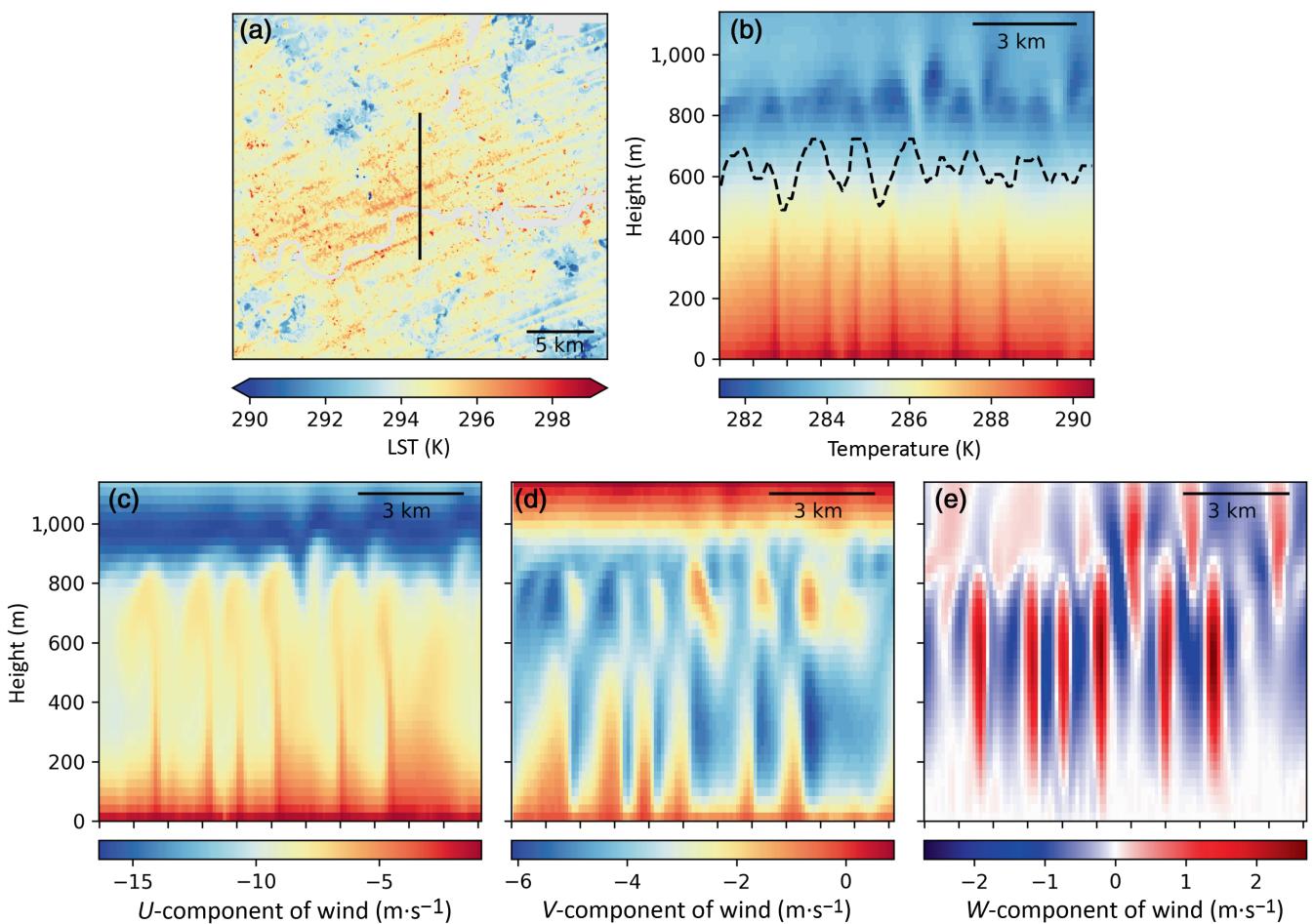


FIGURE 11 UM100 model output for April 22, 2020, at the Landsat observation time (1058 UTC): (a) model domain surface temperature with $\sim 11 \text{ km}$ line showing where the following vertical cross-sections are located; (b) air temperature (dashed line indicates boundary layer height z_i); (c) U , (d) V , and (e) W wind components. Horizontal ticks in (b)–(e) are spaced 1 km apart.

4.5 | Horizontal convective rolls

On April 22, 2020, diagonal stripes are evident in the UM100 but not the Landsat LST (Figure 6), suggesting that the model is unrealistic in this regard. The

stripes are caused by horizontal convective rolls (HCRs) being resolved in the UM100 (Figure 11). HCRs are counter-rotating horizontal turbulent vortices within the convective BL (AMS, 2022; Etling & Brown, 1993; Khanna & Brasseur, 1998; LeMone, 1973; Weckwerth et al., 1999;

Young et al., 2002). HCRs span the depth of the BL (z_i , defined here by the height a moist parcel rising adiabatically becomes negatively buoyant), which is ~ 600 m for the UM100 (Figure 11b) north-south cross-section in the centre of the domain (Figure 11a). The UM100 HCRs have narrow ($\sim 0.5z_i$, warm, positive vertical velocity) updraughts and broad ($\sim 1.5z_i$, cool, negative vertical velocity) downdraughts (Figure 11b,e) parallel to the eastnortheast (246°) mean BL flow. When the updraughts reach the capping inversion, they spread perpendicular to the mean flow (Figure 11c,d), before descending. The updraughts extend throughout the BL down to the surface, causing the high LST stripes. We argue the UM100 having HCRs is correct, but the LST stripes should not be present in the UM100. The latter is supported by the Landsat LST.

There is evidence HCRs are more likely to occur in urban than rural areas due to their high surface roughness (Miao & Chen, 2008). With increasing BL wind shear, convective updraughts have a greater tendency to align with the flow, and for $-z_i/L < 15-20$, HCRs form (Salesky et al., 2017). The Obukhov length L (Garratt, 1992),

$$L = \frac{-u_*^3 T \rho c_p}{\kappa g Q_H}, \quad (4)$$

is a measure of atmospheric stability. The variables are the friction velocity u_* , air temperature T , density of the air ρ , specific heat capacity of air at constant pressure c_p , and the sensible heat flux Q_H ; in addition, κ is the von Kármán constant (0.4) and g is the acceleration due to gravity. These variables are determined at the surface interface with the first level of the model, which is assumed to be at a height equivalent to the roughness length + displacement length + 2 m. For the UM100 we calculate z_i/L to be -4.2 in the region plotted in Figure 11. This value is consistent with there being well-developed HCRs in the UM100.

The UM100 LST stripes are ~ 300 m wide and have $\sim 2^\circ\text{C}$ temperature anomalies (Figure 11a), so it is expected that they would be detected by the ~ 115 m resolution Landsat LST. We therefore conclude that the LST anomalies associated with HCRs in the UM100 are likely unphysical (despite the correct presence of HCRs). Surface layer turbulence is smaller scale than HCR turbulence in the mixed layer, and therefore becomes increasingly less well resolved towards the surface in the UM100 (Figure 11e). High air temperatures associated with the HCR updraughts are extended too far towards the surface by the one-dimensional BL scheme (Boutle et al., 2014; Lock et al., 2000) (Figure 11b), warming the surface. It is likely that this BL turbulence “grey zone” (Honnert

et al., 2020; Wyngaard, 2004) issue would be observed in other $O(100\text{m})$ horizontal grid length NWP models that use one-dimensional BL schemes to parametrise near-surface turbulence.

It is possible that there are <100 m LST anomalies that are not resolved by the UM100 or Landsat. Airborne and tower-based thermal infrared sensors have observed ~ 2 and $0.5-1^\circ\text{C}$ grassland LST anomalies, associated with surface layer streak turbulence (Derksen, 1974; Garai et al., 2013; Schols et al., 1985). For a convective BL with conditions suitable for HCRs ($-z_i/L = 11$), over a grassy field, Garai et al. (2013) estimated LST patterns associated with surface layer streaks to have a streamwise length of 149 m and spanwise width of 69 m. They found that near-surface air temperature is highly correlated with LST and demonstrated that LST patterns move at the wind speed in the upper surface layer. Order 10 m resolution is required for satellite remote sensing to detect LST patterns associated with surface layer streaks. Whether such patterns occur in urban areas is a research question, since urban surface roughness modifies the near-surface turbulence structure (cf. smoother surfaces of the aforementioned studies) (Blunn et al., 2022), and thermal inertia is higher for urban than for vegetated surfaces, meaning that LSTs might respond too slowly to the transient surface-layer turbulence temperature anomalies.

5 | CONCLUSIONS

With the next generation of NWP models moving to higher resolutions ($O(100\text{m})$), new methods to evaluate these models are being explored. In this study we investigate the use of high-resolution thermal imagery from Landsat-8 to assess model performance across Greater London. We conclude that currently there are important benefits to using this data:

- Extensive spatial coverage allows large-scale evaluation of observation–model differences not possible using small area extent observations.
- The high spatial resolution (~ 115 m) means that individual neighbourhoods and land cover variability are well resolved, enabling their representation in hectometric grid length NWP to be evaluated at commensurate scale.
- Spatial patterns visible in the UM100 model data that are not apparent in the Landsat imagery can be identified, allowing model configuration updates and areas for future model development to be identified. These include, in particular:

- i. LST patterns related to the prescription of soil properties and their use in downscaling soil moisture are found on one study day (July 15, 2018). As the NWP model resolution is much higher than before, the consequences not previously considered become apparent. Modifying model configurations, allowed the spatial pattern to be more realistic, but with a larger mean difference.
- ii. Erroneous ~300 m wide stripes identified in the UM100 on one day studied (April 22, 2020) are not evident in the Landsat imagery.

We note certain limitations of using the satellite LST data:

- Single-view satellites sample the surface temperature from a particular angle. For near-nadir field of view observations, near-horizontal facets (e.g., roofs, roads) are heavily weighted while vertical facets are not observed. Ideally, urban canopy modelling would represent the “complete surface” temperature but of the wide variety of approaches taken (e.g., Grimmond et al., 2009, 2011; Lipson et al., 2023), few model all the facets fully because of the computational time being impractical for NWP. Other object resolving models do exist, which can address the surface more fully, but these are computationally orders of magnitude more demanding and/or have limitations of the processes resolved (e.g., radiative transfer only). Hence, there are two definitional differences to be understood: the sampling of the real-world measurement’s field of view and the NWP model simplification of the urban surface. Notably, when a model takes a tile approach the real-world interactions between facets complicates the comparison further as most model grid cells have multiple tile types rather than individual facets (e.g., roofs). While this does not prevent important insights from this comparison, going forward it is critical to reduce definitional uncertainties in model-observation comparison, while improving model skill, to provide a stronger constraint on target behaviour for the model in cities.
- In urban model evaluation, other variables such as surface air temperature and humidity are typically of interest; however, these variables cannot be retrieved from satellite data. Hence, the greatest model benefits may result from combining satellite LST data with other observation sources (e.g., surface energy balance observations).
- Cloud-free conditions are required for satellite LST retrieval, limiting the availability of satellite based LST data for model evaluation.

To address the first of these limitations, and thereby to maximise the value of incorporating urban satellite LST in NWP, both data assimilation developments and model parametrisations are required. Where a model has the necessary internal variables, the preferred approach in data assimilation is to modify the model variables given the constraining observation using an observation operator, which quantifies the observations expected given a model state (e.g., Warren et al., 2018). In the absence of an observation operator, the alternative approach is to attempt to infer from the observed temperature and geometric information the modelled LST. Two preliminary studies provide evidence that this may be helpful. Morrison et al. (2023) demonstrate that, with detailed analysis, the difference between the satellite LST and other representations of temperature can be assessed using three-dimensional radiative transfer modelling, detailed high-resolution surface infrared temperature observations, and very detailed digital models. This approach could provide the basis for an observation operator but is computationally expensive. The Yang et al. (2020) methodology for adjusting the satellite temperature for its geometry is much simpler to deploy, but the empirical parameters currently only exist for one low-latitude city with its specific urban morphology. We have demonstrated the importance of accounting for satellite view angles when using satellite-derived LST data for urban model evaluation. An area needing further research is the development of a fast observation operator that maps between different definitions of temperature given satellite–solar geometry and statistics of the urban landscape.

NWP requires both the model parameters (e.g., albedo, emissivity, land cover, morphology) and the model physics to be correct for successful model performance. This study has provided clear suggestions where model improvement will be beneficial. These relate not just to the model physics but also to the model parameters used. Improved model parameters that are representative of current surface cover, derived using data that resolve the heterogeneous mix of urban surface features appropriately, are essential for assessing the model physics. Improvement of many model parameters (e.g., surface cover data) can also be achieved using other remote sensing data products (e.g., ESA WorldCover) at resolutions appropriate for next-generation NWP and climate modelling.

ACKNOWLEDGMENTS

We thank Claire Bulgin for useful comments on this article.

FUNDING INFORMATION

Research funded by Met Office Strategic Priorities Fund Climate Resilience Programme—100 m-Scale Modelling

for Urban Climate Services: Urban Climate Services Using 100 m Resolution Modelling (IUS-UM100); ERC Synergy Grant urbisphere (855005); NERC EOCIS NE/X019071/1; NERC ASSURE NE/W002965/1.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Processed data will be made available at <https://doi.org/10.5281/zenodo.1003738>.

ORCID

- Thomas W. Hall  <https://orcid.org/0000-0003-0615-3203>
 Lewis Blunn  <https://orcid.org/0000-0002-3207-5002>
 Sue Grimmond  <https://orcid.org/0000-0002-3166-9415>
 Christopher J. Merchant  <https://orcid.org/0000-0003-4687-9850>
 William Morrison  <https://orcid.org/0000-0003-1568-8226>
 Jonathan K. P. Shonk  <https://orcid.org/0000-0003-0479-7590>
 Humphrey Lean  <https://orcid.org/0000-0002-1274-4619>

REFERENCES

- AMS. (2022) Horizontal convective rolls. Glossary of Meteorology. https://glossary.ametsoc.org/wiki/Horizontal_convective_rolls
- Baklanov, A., Grimmond, C., Carlson, D., Terblanche, D., Tang, X., Bouchet, V. et al. (2018) From urban meteorology, climate and environment research to integrated city services. *Urban Climate*, 23, 330–341. Available from: <https://doi.org/10.1016/j.uclim.2017.05.004>
- Berk, A., Anderson, G.P., Bernstein, L.S., Acharya, P.K., Dothe, H., Matthew, M.W. et al. (1999) MODTRAN4 radiative transfer modeling for atmospheric correction. Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE). 3rd Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research, Denver, Co.
- Best, M.J. (2005) Representing urban areas within operational numerical weather prediction models. *Boundary-Layer Meteorology*, 114(1), 91–109. Available from: <https://doi.org/10.1007/s10546-004-4834-5>
- Best, M.J., Pryor, M., Clark, D.B., Rooney, G.G., Essery, R.L.H., Ménard, C.B. et al. (2011) The joint UK land environment simulator (JULES), model description – part 1: energy and water fluxes. *Geoscientific Model Development*, 4(3), 677–699. Available from: <https://doi.org/10.5194/gmd-4-677-2011>
- Blunn, L.P. (2021) *Characterising mixing and pollution transport in the urban boundary layer*. PhD thesis. Reading, UK: University of Reading.
- Blunn, L.P., Coceal, O., Nazarian, N., Barlow, J.F., Plant, R.S., Bohnenstengel, S.I. et al. (2022) Turbulence characteristics across a range of idealized urban canopy geometries. *Boundary-Layer Meteorology*, 182(2), 275–307. Available from: <https://doi.org/10.1007/s10546-021-00658-6>
- Bohenstengel, S.I., Evans, S., Clark, P.A. & Belcher, S.E. (2011) Simulations of the London urban heat Island. *Quarterly Journal of the Royal Meteorological Society*, 137(659), 1625–1640. Available from: <https://doi.org/10.1002/qj.855>
- Bohenstengel, S.I. & Hendry, M.A. (2016) *Report on implementation and evaluation of MORUSES in the UKV (PS37)*. Exeter, UK: Met Office. <https://digital.nmla.metoffice.gov.uk/IO&uscore;30a48cac-662d-48ce-9a63-f5e81496c665> [Accessed 5th February 2023].
- Boutle, I.A., Eyre, J.E.J. & Lock, A.P. (2014) Seamless stratocumulus simulation across the turbulent gray zone. *Monthly Weather Review*, 142(4), 1655–1668. Available from: <https://doi.org/10.1175/MWR-D-13-00229.1>
- Boutle, I.A., Finnenkoetter, A., Lock, A.P. & Wells, H. (2016) The London model: forecasting fog at 333 m resolution. *Quarterly Journal of the Royal Meteorological Society*, 142(694), 360–371. Available from: <https://doi.org/10.1002/qj.2656>
- Bunce, R.G.H., Barr, C.J., Clarke, R.T., Howard, D.C. & Lane, A.M.J. (1990) ITE land classification of Great Britain 1990. Available from: <https://doi.org/10.5285/ab320e08-faf5-48e1-9ec9-77a213d2907f>
- Bush, M., Boutle, I., Edwards, J., Finnenkoetter, A., Franklin, C., Hanley, K. et al. (2022) The second met Office unified model/JULES regional atmosphere and land configuration, RAL2. *Geoscientific Model Development Discussion*, 2022, 1–35. Available from: <https://doi.org/10.5194/gmd-2022-209>
- Bush, M., Flack, D., Arnold, A., Best, M., Bohnenstengel, S., Boutle, I. et al. (2024). Unifying mid-latitude and tropical regional model configurations: the third met Office unified model–JULES regional atmosphere and land configuration, RAL3. In prep.
- Chapman, L., Bell, C. & Bell, S. (2017) Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat Island of London quantified using Netatmo weather stations. *International Journal of Climatology*, 37(9), 3597–3605. Available from: <https://doi.org/10.1002/joc.4940>
- Clark, D.B., Mercado, L.M., Sitch, S., Jones, C.D., Gedney, N., Best, M.J. 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. Available from: <https://doi.org/10.5194/gmd-4-701-2011>
- Coney, J., Pickering, B., Dufton, D., Lukach, M., Brooks, B. & Neely, R.R., III. (2022) How useful are crowdsourced air temperature observations? An assessment of Netatmo stations and quality control schemes over the United Kingdom. *Meteorological Applications*, 29(3), e2075. Available from: <https://doi.org/10.1002/met.2075>
- Cook, M., Schott, J.R., Mandel, J. & Raqueno, N. (2014) Development of an operational calibration methodology for the Landsat thermal data archive and initial testing of the atmospheric compensation component of a land surface temperature (LST) product from the archive. *Remote Sensing*, 6(11), 11244–11266. Available from: <https://doi.org/10.3390/rs61111244>
- Cornes, R.C., Dirksen, M. & Sluiter, R. (2019) Correcting citizen-science air temperature measurements across The

- Netherlands for short wave radiation bias. *Meteorological Applications*, 27(1), e1814. Available from: <https://doi.org/10.1002/met.1814>
- Davies, T., Cullen, M.J.P., Malcolm, A.J., Mawson, M.H., Staniforth, A., White, A.A. et al. (2005) A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 131(608), 1759–1782. Available from: <https://doi.org/10.1256/qj.04.101>
- Derksen, W.J. (1974) Thermal infrared pictures and the mapping of microclimate. *Netherlands Journal of Agricultural Science*, 22, 119–132.
- DUKES. (2003) *Digest of United Kingdom energy statistics 2003*. London, UK: Department of Trade and Industry's Energy Group, The Stationery Office.
- Etling, D. & Brown, R.A. (1993) Roll vortices in the planetary boundary layer: a review. *Boundary-Layer Meteorology*, 65(3), 215–248. Available from: <https://doi.org/10.1007/BF00705527>
- Fenner, D., Bechtel, B., Demuzere, M., Kittner, J. & Meier, F. (2021) CrowdQC+—a quality-control for crowdsourced air-temperature observations enabling world-wide urban climate applications. *Frontiers in Environmental Science*, 9, 720747. Available from: <https://doi.org/10.3389/fenvs.2021.720747>
- Fenner, D., Christen, A., Grimmond, S., Meier, F., Morrison, W. & Zeeman, M. (2022) Chasing the plume – Investigating the urban boundary layer (downwind) of a city with transects of fixed and mobile ceilometer measurement. 2022 IAUC Virtual Poster Conference. Available from: <https://iaucposter2022.com/wp-content/uploads/2022/08/AbstractBook&uscore;2022IAUCposter.pdf> [Accessed 5th November 2023].
- Field, P.R., Hill, A., Shipway, B., Furtado, K., Wilkinson, J., Miltenberger, A. et al. (2023) Implementation of a double moment cloud microphysics scheme in the UK met office regional numerical weather prediction model. *Quarterly Journal of the Royal Meteorological Society*, 149(752), 703–739. Available from: <https://doi.org/10.1002/qj.4414>
- Fuller, R., Groom, G.B. & Jones, A.R.C. (1994) The land cover map of Great Britain: an automated classification of Landsat thematic mapper data. *Photogrammetric Engineering and Remote Sensing*, 60, 553–562.
- Garai, A., Pardyjak, E., Steeneveld, G.-J. & Kleissl, J. (2013) Surface temperature and surface-layer turbulence in a convective boundary layer. *Boundary-Layer Meteorology*, 148(1), 51–72. Available from: <https://doi.org/10.1007/s10546-013-9803-4>
- Garratt, J.R. (1992) Scaling laws for mean and turbulent quantities. In: *The atmospheric boundary layer*. Cambridge, UK: Cambridge University Press, p. 38.
- Gerace, A. & Montanaro, M. (2017) Derivation and validation of the stray light correction algorithm for the thermal infrared sensor onboard Landsat 8. *Remote Sensing of Environment*, 191, 246–257. Available from: <https://doi.org/10.1016/j.rse.2017.01.029>
- Grimmond, C.S.B., Best, M., Barlow, J., Arnfield, A.J., Baik, J.-J., Belcher, S. et al. (2009) Urban surface energy balance models: model characteristics & methodology for a comparison study. In: Baklanov, A., Grimmond, C.S.B., Mahura, A. & Athanassiadou, M. (Eds.) *Meteorological and Air Quality Models for Urban Areas*. Heidelberg: Springer-Verlag. Available from: <https://doi.org/10.1007/978-3-642-00298-4&uscore;11>
- Grimmond, C.S.B., Blackett, M., Best, M.J., Baik, J.-J., Belcher, S.E., Beringer, J. et al. (2011) Initial results from phase 2 of the international urban energy balance model comparison. *International Journal of Climatology*, 31, 244–272. Available from: <https://doi.org/10.1002/joc.2227>
- Grimmond, S. & Ward, H.C. (2021) Urban measurements and their interpretation. In: Foken, T. (Ed.) *Springer Handbook of Atmospheric Measurements*. Cham: Springer, pp. 1391–1423. Available from: <https://doi.org/10.1007/978-3-030-52171-4&uscore;52>
- Hallsworth, C. (2007) *Unified model documentation paper 28: sub-grid turbulence scheme*. Exeter, UK: Met Office. Available from: <https://code.metoffice.gov.uk/doc/um/latest/papers/umdp&uscore;028.pdf> [Accessed 5th February 2023]
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. Available from: <https://doi.org/10.1002/qj.3803>
- Hertwig, D., Grimmond, C., Kotthaus, S., Vanderwel, C., Gough, H., Haefelin, M. et al. (2021) Variability of physical meteorology in urban areas at different scales: implications for air quality. *Faraday Discussions*, 226, 149–172. Available from: <https://doi.org/10.1039/D0FD00098A>
- Hertwig, D., Grimmond, S., Hendry, M.A., Saunders, B., Wang, Z., Jeoffrion, M. et al. (2020) Urban signals in high-resolution weather and climate simulations: role of urban land-surface characterisation. *Theoretical and Applied Climatology*, 142(1), 701–728. Available from: <https://doi.org/10.1007/s00704-020-03294-1>
- Hollmann, R., Merchant, C.M., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A. et al. (2013) The ESA climate change initiative: satellite data records for Essential Climate Variables. *Bulletin of the American Meteorological Society*, 94(10), 1541–1552. Available from: <https://doi.org/10.1175/BAMS-D-11-00254.1>
- Honnert, R., Efthathiou, G.A., Beare, R.J., Ito, J., Lock, A., Neggers, R. et al. (2020) The atmospheric boundary layer and the “gray zone” of turbulence: a critical review. *Journal of Geophysical Research: Atmospheres*, 125(13), e2019JD030317. Available from: <https://doi.org/10.1029/2019JD030317>
- Hulley, G., Hook, S., Abbott, E., Malakar, N., Islam, T. & Abrams, M. (2015) The ASTER global emissivity dataset (ASTER GED): mapping Earth's emissivity at 100 meter spatial scale. *Geophysical Research Letters*, 42(19), 7966–7976. Available from: <https://doi.org/10.1002/2015gl065564>
- Jiang, L., Zhan, W., Voogt, J., Zhao, L., Gao, L., Huang, F. et al. (2018) Remote estimation of complete urban surface temperature using only directional radiometric temperatures. *Building and Environment*, 135, 224–236. Available from: <https://doi.org/10.1016/j.buildenv.2018.03.005>
- Jimenez-Munoz, J.C., Cristobal, J., Sobrino, J.A., Soria, G., Ninyerola, M. & Pons, X. (2009) Revision of the Single-Channel algorithm for land surface temperature retrieval from Landsat thermal-infrared data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1), 339–349. Available from: <https://doi.org/10.1109/tgrs.2008.2007125>
- Jimenez-Munoz, J.C., Sobrino, J.A., Skokovic, D., Mattar, C. & Cristobal, J. (2014) Land surface temperature retrieval methods from Landsat-8 thermal infrared sensor data. *IEEE Geoscience and Remote Sensing Letters*, 11(10), 1840–1843. Available from: <https://doi.org/10.1109/lgrs.2014.2312032>
- Joe, P., Belair, S., Bernier, N.B., Bouchet, V., Brook, J.R., Brunet, D. et al. (2018) The Environment Canada Pan and Parapan

- American science showcase project. *Bulletin of the American Meteorological Society*, 99(5), 921–953. Available from: <https://doi.org/10.1175/bams-d-16-0162.1>
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L. et al. (1996) The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3), 437–471. Available from: [https://doi.org/10.1175/1520-0477\(1996\)077<0437:Tnyrp>2.0.Co;2](https://doi.org/10.1175/1520-0477(1996)077<0437:Tnyrp>2.0.Co;2)
- Khanna, S. & Brasseur, J.G. (1998) Three-dimensional buoyancy- and shear-induced local structure of the atmospheric boundary layer. *Journal of the Atmospheric Sciences*, 55(5), 710–743. Available from: [https://doi.org/10.1175/1520-0469\(1998\)055<0710:TDBASI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<0710:TDBASI>2.0.CO;2)
- Kirk, P.J., Clark, M.R. & Creed, E. (2021) Weather observations website. *Weather*, 76(2), 47–49. Available from: <https://doi.org/10.1002/wea.3856>
- Kottek, M., Grieser, J., Beck, C., Rudolf, B. & Rubel, F. (2006) World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3), 259–263. Available from: <https://doi.org/10.1127/0941-2948/2006/0130>
- Kotthaus, S., Haeffelin, M., Bastin, S., Boissard, C., Bouffies-Cloché, S., Canut, G. et al. (2023) PANAME – project synergy of atmospheric research in the Paris region. 11th International Conference on Urban Climate. Australia: UNSW Sydney.
- Lean, H.W., Barlow, J.F. & Halios, C.H. (2019) The impact of spin-up and resolution on the representation of a clear convective boundary layer over London in order 100 m grid-length versions of the Met Office Unified Model. *Quarterly Journal of the Royal Meteorological Society*, 145(721), 1674–1689. Available from: <https://doi.org/10.1002/qj.3519>
- LeMone, M.A. (1973) The structure and dynamics of horizontal roll vortices in the planetary boundary layer. *Journal of Atmospheric Sciences*, 30(6), 1077–1091. Available from: [https://doi.org/10.1175/1520-0469\(1973\)030<1077:TSADOH>2.0.CO;2](https://doi.org/10.1175/1520-0469(1973)030<1077:TSADOH>2.0.CO;2)
- Leroyer, S., Béclair, S., Mailhot, J. & Strachan, I.B. (2011) Microscale numerical prediction over Montreal with the Canadian external urban modeling system. *Journal of Applied Meteorology and Climatology*, 50, 2410–2428. Available from: <https://doi.org/10.1175/JAMC-D-11-013.1>
- Lilly, D.K. (1962) On the numerical simulation of buoyant convection. *Tellus*, 14(2), 148–172. Available from: <https://doi.org/10.3402/tellusa.v14i2.9537>
- Lindberg, F. & Grimmond, C.S.B. (2011a) The influence of vegetation and building morphology on shadow patterns and mean radiant temperatures in urban areas: model development and evaluation. *Theoretical and Applied Climatology*, 105(3), 311–323. Available from: <https://doi.org/10.1007/s00704-010-0382-8>
- Lindberg, F. & Grimmond, C.S.B. (2011b) Nature of vegetation and building morphology characteristics across a city: influence on shadow patterns and mean radiant temperatures in London. *Urban Ecosystems*, 14(4), 617–634. Available from: <https://doi.org/10.1007/s11252-011-0184-5>
- Lindberg, F., Grimmond, C.S.B., Gabey, A., Huang, B., Kent, C.W., Sun, T. et al. (2018) Urban multi-scale environmental predictor (UMEP): an integrated tool for city-based climate services. *Environmental Modelling & Software*, 99, 70–87. Available from: <https://doi.org/10.1016/j.envsoft.2017.09.020>
- Lipson, M.J., Grimmond, S., Best, M., Abramowitz, G., Coutts, A., Tapper, N. et al. (2023) Evaluation of 30 urban land surface models in the urban-PLUMBER project: phase 1 results. *Quarterly Journal of the Royal Meteorological Society*, 150(758), 126–169. Available from: <https://doi.org/10.1002/qj.4589>
- Lock, A.P., Brown, A.R., Bush, M.R., Martin, G.M. & Smith, R.N.B. (2000) A new boundary layer mixing scheme. Part I: scheme description and single-column model tests. *Monthly Weather Review*, 128(9), 3187–3199. Available from: [https://doi.org/10.1175/1520-0493\(2000\)128<3187:ANBLMS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<3187:ANBLMS>2.0.CO;2)
- Lucchesi, R. (2013) File specification for GEOS-5 FP-IT. GMAO Office Note No. 2 (version 1.2). <https://ntrs.nasa.gov/citations/20150001438> [Accessed 5th November 2023].
- Malakar, N.K., Hulley, G.C., Hook, S.J., Laraby, K., Cook, M. & Schott, J.R. (2018) An operational land surface temperature product for Landsat thermal data: methodology and validation. *IEEE Transactions on Geoscience and Remote Sensing*, 56(10), 5717–5735. Available from: <https://doi.org/10.1109/TGRS.2018.2824828>
- Masson, V., Grimmond, C.S.B. & Oke, T.R. (2002) Evaluation of the town energy balance (TEB) scheme with direct measurements from dry districts in two cities. *Journal of Applied Meteorology*, 41(10), 1011–1026. Available from: [https://doi.org/10.1175/1520-0450\(2002\)041<1011:Eottb>2.0.Co;2](https://doi.org/10.1175/1520-0450(2002)041<1011:Eottb>2.0.Co;2)
- Met Office. (2012) *Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current)*. Oxford, UK: NCAS British Atmospheric Data Centre. <http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0> [Accessed 18th August 2023].
- Miao, S.G. & Chen, F. (2008) Formation of horizontal convective rolls in urban areas. *Atmospheric Research*, 89(3), 298–304. Available from: <https://doi.org/10.1016/j.atmosres.2008.02.013>
- Morrison, W., Grimmond, S. & Kotthaus, S. (2023) Simulating satellite urban land surface temperatures: sensitivity to sensor view angle and assumed landscape complexity. *Remote Sensing of Environment*, 293, 113579. Available from: <https://doi.org/10.1016/j.rse.2023.113579>
- Morrison, W., Kotthaus, S. & Grimmond, S. (2021) Urban surface temperature observations from ground-based thermography: intra- and inter-facet variability. *Urban Climate*, 35, 100748. Available from: <https://doi.org/10.1016/j.juclim.2020.100748>
- Muller, C.L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G. et al. (2015) Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology*, 35(11), 3185–3203. Available from: <https://doi.org/10.1002/joc.4210>
- Parastatidis, D., Mitraka, Z., Chrysoulakis, N. & Abrams, M. (2017) Online global land surface temperature estimation from Landsat. *Remote Sensing*, 9(12), 1208. Available from: <https://doi.org/10.3390/rs9121208>
- Porson, A., Clark, P.A., Harman, I.N., Best, M.J. & Belcher, S.E. (2010a) 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(651), 1514–1529. Available from: <https://doi.org/10.1002/qj.668>
- Porson, A., Clark, P.A., Harman, I.N., Best, M.J. & Belcher, S.E. (2010b) Implementation of a new urban energy budget scheme into MetUM. Part II: validation against observations and model intercomparison. *Quarterly Journal of the Royal Meteorological Society*, 136(651), 1530–1542. Available from: <https://doi.org/10.1002/qj.572>
- Salesky, S.T., Chamecki, M. & Bou-Zeid, E. (2017) On the nature of the transition between roll and cellular Organization in

- the Convective Boundary Layer. *Boundary-Layer Meteorology*, 163(1), 41–68. Available from: <https://doi.org/10.1007/s10546-016-0220-3>
- Saunders, B., Grimmond, S., Hertwig, D., Bohnenstengel, S.I., Lean, H. & Morrison, W. (2024) Considerations for using large aperture scintillometry sensible heat fluxes to dynamically evaluate numerical weather predictions. *Quarterly Journal of the Royal Meteorological Society* (In Review).
- Saunders, R., Hocking, J., Turner, E., Havemann, S., Geer, A., Lupu, C. et al. (2020) RTTOV-13 Science and Validation Report. EUMETSAT NWP SAF. <https://nwp-saf.eumetsat.int/site/download/documentation/rtm/docs&uscore;rttov13/rttov13&uscore;svr.pdf> [Accessed 5th November 2023].
- Schols, J.L.J., Jansen, A.E. & Krom, J.G. (1985) Characteristics of turbulent structures in the unstable atmospheric surface layer. *Boundary-Layer Meteorology*, 33(2), 173–196. Available from: <https://doi.org/10.1007/BF00123389>
- Smagorinsky, J. (1963) General circulation experiments with the primitive equations: I. The basic experiment. *Monthly Weather Review*, 91(3), 99–164. Available from: [https://doi.org/10.1175/1520-0493\(1963\)091<0099:Gcewtp>2.3.Co;2](https://doi.org/10.1175/1520-0493(1963)091<0099:Gcewtp>2.3.Co;2)
- Tang, Y., Sun, T., Luo, Z., Omidvar, H., Theeuwes, N., Xie, X. et al. (2021) Urban meteorological forcing data for building energy simulations. *Building and Environment*, 204, 108088. Available from: <https://doi.org/10.1016/j.buildenv.2021.108088>
- Theeuwes, N.E., Ronda, R.J., Harman, I.N., Christen, A. & Grimmond, C.S.B. (2019) Parametrizing horizontally-averaged wind and temperature profiles in the urban roughness sublayer. *Boundary-Layer Meteorology*, 173(3), 321–348. Available from: <https://doi.org/10.1007/s10546-019-00472-1>
- USGS. (2019a) Landsat 8 (L8) Data Users Handbook Version 5.0. Department of the Interior U.S. Geological Survey. <https://www.usgs.gov/media/files/landsat-8-data-users-handbook> [Accessed 5th November 2023]
- USGS. (2019b) Landsat 8 Prelaunch Point Spread Functions. <https://www.usgs.gov/media/files/landsat-8-prelaunch-point-spread-functions> [Accessed 5th November 2023].
- USGS. (2021a) Landsat 8. <https://www.usgs.gov/landsat-missions/landsat-8> [Accessed 5th November 2023].
- USGS. (2021b) Landsat collection 2 Level-1 data. <https://www.usgs.gov/landsat-missions/landsat-collection-2-level-1-data> [Accessed 5th November 2023].
- Van Weverberg, K., Morcrette, C.J. & Boutle, I. (2021b) A bimodal diagnostic cloud fraction parameterization. Part II: evaluation and resolution sensitivity. *Monthly Weather Review*, 149(3), 859–878. Available from: <https://doi.org/10.1175/MWR-D-20-0230.1>
- Van Weverberg, K., Morcrette, C.J., Boutle, I., Furtado, K. & Field, P.R. (2021a) A bimodal diagnostic cloud fraction parameterization. Part I: motivating analysis and scheme description. *Monthly Weather Review*, 149(3), 841–857. Available from: <https://doi.org/10.1175/MWR-D-20-0224.1>
- Vesala, T., Järvi, L., Launiainen, S., Sogachev, A., Rannik, Ü., Mammarella, I. et al. (2008) Surface–atmosphere interactions over complex urban terrain in Helsinki, Finland. *Tellus B: Chemical and Physical Meteorology*, 60(2), 188–199. Available from: <https://doi.org/10.1111/j.1600-0889.2007.00312.x>
- Voogt, J.A. (2000) Image representations of complete urban surface temperatures. *Geocarto International*, 15(3), 21–32. Available from: <https://doi.org/10.1080/10106040008542160>
- Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S. et al. (2017) The Met Office unified model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations. *Geoscientific Model Development*, 10(4), 1487–1520. Available from: <https://doi.org/10.5194/gmd-10-1487-2017>
- Wan, Z. & Dozier, J. (1996) A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, 34(4), 892–905. Available from: <https://doi.org/10.1109/36.508406>
- Wan, Z., Hook, S. & Hulley, G. (2021) MODIS/Terra land surface temperature/emissivity daily L3 global 1km SIN grid V061 [Dataset]. NASA EOSDIS Land Processes Distributed Active Archive Center. Available from: <https://doi.org/10.5067/MODIS/MOD11A1.061> [Accessed 5th November 2023]
- Wang, D., Chen, Y., Hu, L. & Voogt, J.A. (2022) Urban thermal anisotropy: a comparison among observational and modeling approaches at different time scales. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–15. Available from: <https://doi.org/10.1109/TGRS.2022.3171285>
- Wang, W. (2014) Analytically modelling mean wind and stress profiles in canopies. *Boundary-Layer Meteorology*, 151(2), 239–256. Available from: <https://doi.org/10.1007/s10546-013-9899-6>
- Ward, H.C., Evans, J.G. & Grimmond, C.S.B. (2014) Multi-scale sensible heat fluxes in the suburban environment from large-aperture Scintillometry and Eddy covariance. *Boundary-Layer Meteorology*, 152(1), 65–89. Available from: <https://doi.org/10.1007/s10546-014-9916-4>
- Warren, E., Charlton-Perez, C., Kotthaus, S., Lean, H., Ballard, S., Hopkin, E. et al. (2018) Evaluation of forward-modelled attenuated backscatter using an urban ceilometer network in London under clear-sky conditions. *Atmospheric Environment*, 191, 532–547. Available from: <https://doi.org/10.1016/j.atmosenv.2018.04.045>
- Weckwerth, T.M., Horst, T.W. & Wilson, J.W. (1999) An observational study of the evolution of horizontal convective rolls. *Monthly Weather Review*, 127(9), 2160–2179. Available from: [https://doi.org/10.1175/1520-0493\(1999\)127<2160:AOSOTE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2160:AOSOTE>2.0.CO;2)
- WMO. (2018a) Guide to instruments and methods of observation, volume I – measurement of meteorological variables (WMO-No. 8). https://library.wmo.int/viewer/41650/download?file=8_I_2021&uscore;en.pdf&type=pdf&navigator=1 [Accessed 24th October 2023].
- WMO. (2018b) Guide to instruments and methods of observation, volume III – observing systems (WMO-No. 8). https://library.wmo.int/viewer/41650/download?file=8_III_2021_en.pdf&type=pdf&navigator=1 [Accessed 24th October 2023].
- WMO. (2019) Guidance on integrated urban hydrometeorological, climate and environmental services, volume I: concept and methodology (WMO-No. 1234). <https://library.wmo.int/idurl/4/56760> [Accessed 26th October 2023].
- Wood, N., Staniforth, A., White, A., Allen, T., Diamantakis, M., Gross, M. et al. (2014) An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quarterly Journal of the Royal Meteorological Society*, 140(682), 1505–1520. Available from: <https://doi.org/10.1002/qj.2235>

Wyngaard, J.C. (2004) Toward numerical modeling in the “Terra incognita”. *Journal of the Atmospheric Sciences*, 61(14), 1816–1826. Available from: [https://doi.org/10.1175/1520-0469\(2004\)061<1816:TNTMITT>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<1816:TNTMITT>2.0.CO;2)

Yang, J., Wong, M.S., Ho, H.C., Krayenhoff, E.S., Chan, P.W., Abbas, S. et al. (2020) A semi-empirical method for estimating complete surface temperature from radiometric surface temperature, a study in Hong Kong city. *Remote Sensing of Environment*, 237, 111540. Available from: <https://doi.org/10.1016/j.rse.2019.111540>

Young, G.S., Kristovich, D.A.R., Hjelmfelt, M.R. & Foster, R.C. (2002) Rolls, streets, waves, and more: a review of quasi-two-dimensional structures in the atmospheric boundary layer. *Bulletin of the American Meteorological Society*, 83(7), 997–1002. Available from: [https://doi.org/10.1175/1520-0477\(2002\)083<0997:RSWAMA>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0997:RSWAMA>2.3.CO;2)

SUPPORTING INFORMATION

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

How to cite this article: Hall, T.W., Blunn, L., Grimmond, S., McCarroll, N., Merchant, C.J., Morrison, W. *et al.* (2024) Utility of thermal remote sensing for evaluation of a high-resolution weather model in a city. *Quarterly Journal of the Royal Meteorological Society*, 150(760), 1771–1790. Available from: <https://doi.org/10.1002/qj.4669>