Methodology to evaluate numerical weather predictions using large-aperture scintillometry sensible heat fluxes: demonstration in London

Received: 2 April 2024 | Revised: 12 July 2024 | Accepted: 26 July 2024

DOI: 10.1002/qj.4837

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Methodology to evaluate numerical weather predictions using large-aperture scintillometry sensible heat fluxes: Demonstration in London

Beth Saunders[1] | **Sue Grimmond**[1] | **Denise Hertwig**[1] | **Sylvia I. Bohnenstengel**[2] | **Humphrey W. Lean**[2] | **William Morrison**[1]

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

[2]MetOffice@Reading, Reading, UK

Correspondence
Sue Grimmond, Department of Meteorology, University of Reading, Earley Gate, Reading, RG6 6ET, UK.
Email: c.s.grimmond@reading.ac.uk

Funding information
NERC Case studentship, Grant/Award Number: NE/N008103/1; ERC urbisphere, Grant/Award Number: 855005

Abstract

To evaluate numerical weather predictions (NWP) using observations, the horizontal, vertical and temporal resolution characteristics need to be considered to ensure consistency. Here, large-aperture scintillometry (LAS) is used to derive turbulent sensible heat fluxes (Q_H), as the source area extents are comparable to NWP with horizontal grid box resolutions of order 1 km. We demonstrate our methodology using LAS observations undertaken in central London and the Met Office's operational forecast model for the United Kingdom (UKV). In the horizontal, we ensure consistency between the LAS source area locations and extent, as they vary with wind direction and stability, and therefore the NWP grid boxes that should be selected for comparison. In the vertical, the appropriate model level needs to be selected relative to the changing effective measurement height (z_f) of the observations. The LAS fluxes time and space averaging allows Q_H to be obtained for time periods similar to the model timestep (e.g., 1 min) but the UKV does not explicitly represent higher-frequency turbulent motions. During two spring days, three to five UKV grid boxes (1.5 km resolution) fall within the LAS source area (60%). However, with the model land cover data lacking realistic spatial variability, modelled effective surface Q_H spatial variability is small, leaving a central grid box on the LAS path representative. Larger differences occur between modelled Q_H at the surface and closest to z_f, despite both theoretically being within the inertial sublayer. The modelled Q_H has better agreement with observed fluxes when averaged over long periods (10+ min *cf.* 1 min). As LAS-derived Q_H has larger source areas than eddy covariance (EC) measurements, there are benefits to assessing NWP model performance with LAS. With multiple paths, spatial variability in response to changes in surface cover could be assessed at shorter times scales than possible with EC.

KEYWORDS

large-aperture scintillometry, London, model evaluation, numerical weather prediction, sensible heat flux, source area, urban

1 | INTRODUCTION

Increasing spatial resolutions of numerical weather prediction (NWP) models, together with the use of suitable urban land surface schemes, make it possible to capture spatial variations across neighbourhoods and their impact on the near-surface atmosphere (Barlow et al., 2017). State-of-the-art operational regional NWP models have grid lengths of the order of 1 km (e.g., 1.5 km, Met Office's UKV – Tang et al., 2013; ~2 km, DWD's ICON-D2 –Giorgetta et al., 2018; Reinert et al., 2021). However, the next generation of NWP models being developed have grid lengths of the order of 100 m (i.e., hectare), or O(100 m) using the Landau symbol, as enhanced high-performance computing facilities become available (Lean et al., 2019; Leroyer et al., 2014; Ronda et al., 2017).

With large proportions of both the English (>83%, GOV.UK, 2023) and global population (United Nations, 2019) living in urban areas, the ability to accurately forecast weather conditions in cities, for routine and hazardous or extreme conditions (e.g., heatwaves) exacerbated by climate change (Kendon et al., 2014) is critical. Between neighbourhoods in close proximity (e.g., central business district, parks, different residential areas) both urban form (e.g., building morphology, land cover, urban fabric) and function (e.g., anthropogenic heat emissions) vary, influencing urban surface–atmosphere exchanges (Britter & Hanna, 2003; Stewart & Oke, 2012) creating distinct local climates (Oke et al., 2017).

As operational NWP models O(1 km) do not resolve buildings explicitly (Best, 2005), they need to be parametrised (Masson, 2006) by an urban land surface model (ULSM) accounting for the aerodynamic, radiative and thermal properties and anthropogenic heat emissions by calculating the surface energy balance fluxes (Oke, 1988) for the top of individual grid boxes:

$$Q^* + Q_F = \Delta Q_S + Q_H + Q_E, \quad (1)$$

where Q^* is the net all-wave radiation which is a function of the incoming (↓) and outgoing (↑) short-wave (K) and long-wave (L) radiation (Oke et al., 2017):

$$Q^* = \left(K_\downarrow - K_\uparrow\right) + \left(L_\downarrow - L_\uparrow\right). \quad (2)$$

Q_F is the anthropogenic heat flux, ΔQ_S is the net storage heat flux, Q_H is the turbulent sensible heat flux, and Q_E is the turbulent latent heat flux (all W m^{-2}). As subgrid-scale (SGS) processes are parametrised within a grid box, SGS advection is not explicit, whilst inter grid box advection is explicitly simulated by the NWP model.

Each term in Equation (1), and the urban water balance (Grimmond et al., 1986), are governed by the surface characteristics (e.g., surface radiative properties of albedo and emissivity). These in turn modify other atmospheric variables (e.g., air temperatures, wind, humidity). Hence, flux partitioning into Q_H or Q_E varies with surface characteristics both between cities (e.g., Grimmond & Oke, 2002) and within cities (e.g., Offerle et al., 2006). ΔQ_S accounts for the heating and cooling of the urban volume including the ground (Oke et al., 2017). An important difference from non-urban areas is the Q_F term, associated with human activities (e.g., Capel-Timms et al., 2020), but varying with socio-economic and geographic settings (Allen et al., 2011).

Turbulent heat fluxes have been observed in many locations around the world using in-situ eddy covariance (EC) measurement techniques (as summarised by Kotthaus & Grimmond, 2014a, 2014b). A sonic anemometer with infrared gas analyser needs to be mounted high above the roughness elements (e.g., on a mast) to be above the roughness sublayer (RSL, Figure 1a). EC observations have been extensively used to evaluate urban land surface models, for example, Town Energy Balance (TEB) model (Lemonsu et al., 2004); Community Land Model (CLM) (Karsisto et al., 2016); Surface Urban Energy and Water balance Scheme (SUEWS) (Ward et al., 2016); Joint UK Land Environment Simulator (JULES) 1 Tile (1 T) (Best et al., 2006); JULES Met Office–Reading Urban Surface Exchange Scheme (MORUSES) (Hertwig et al., 2020).

Another technique to observe turbulent heat fluxes over cities is large-aperture scintillometry (LAS; e.g., Crawford et al., 2017; Lagouarde et al., 2006; Ward et al., 2014). Although primarily used for measuring Q_H (Ward, 2017), Q_E is obtainable if combined with a millimetre-wave scintillometer (Ward et al., 2015a, 2015b), or by prescribing an assumed Bowen ratio value(s) (Ward et al., 2014). In many environments, EC sensors have been used to demonstrate LAS performance (Beyrich et al., 2021), which in urban areas includes central city (e.g., Marseille: Lagouarde et al., 2006) and residential areas (e.g., Swindon: Ward et al., 2014).

Large-aperture scintillometry requires a clear atmospheric path (free of obstacles) between a transmitter and receiver located O (0.5–5 km) apart (Ward, 2017). Like EC techniques, LAS measurements need to be undertaken in the inertial sublayer (ISL, Figure 1a). As the most influential part of the LAS weighting function is in the centre of the path (Ward, 2017) (Figure 1c), the LAS transmitter and receiver can be deployed on tall structures without massively influencing the measured flow fields, making deployment less challenging (Grimmond & Ward, 2021) than EC siting in urban areas (e.g., Barlow et al., 2011). Also, the LAS source area extent (SA_{LAS}) is more comparable to NWP grid lengths than the EC counterpart (Figure 1c, Section 2.1).

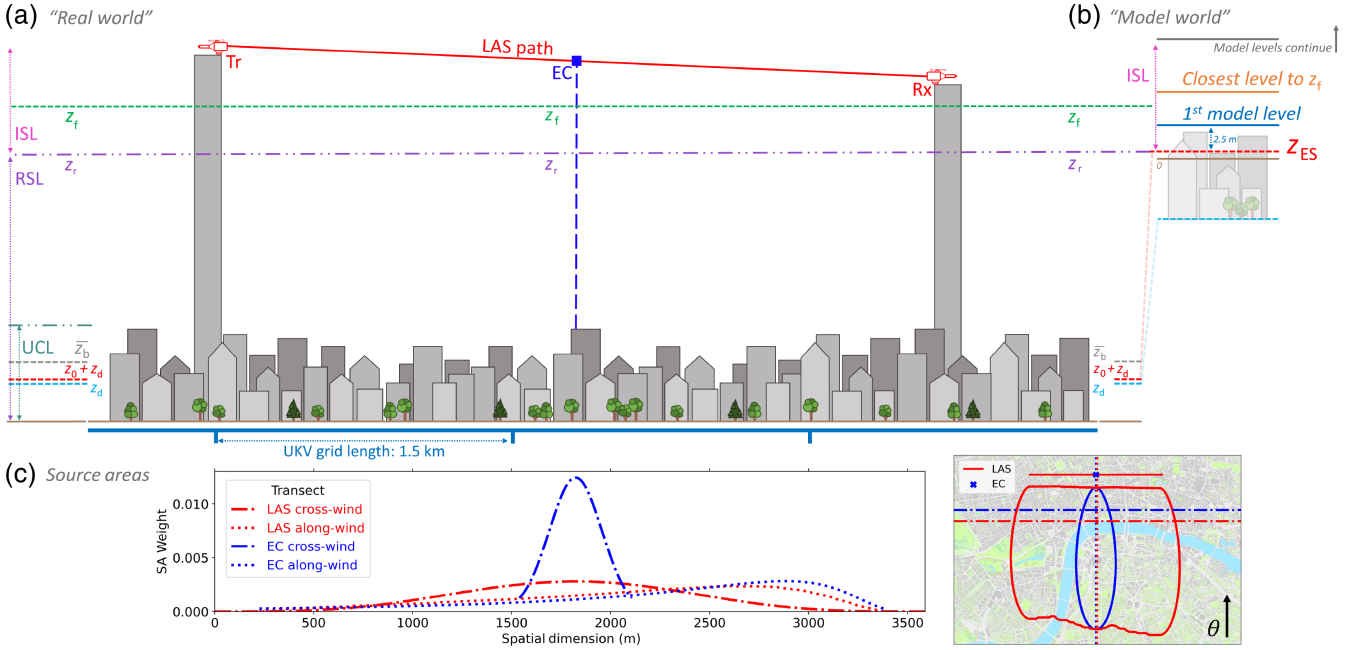

FIGURE 1 Schematic representation of a large-aperture scintillometer (LAS) path between a transmitter (Tr) and receiver (Rx) mounted on two buildings significantly taller than the mean building height ($\overline{z_b}$). Their locations (not to scale) are relative to the urban canopy layer (UCL), roughness sublayer (RSL), blending height (z_r), inertial sublayer (ISL) or constant flux layer, and the sum of displacement height z_d and momentum roughness length z_0 in the: (a) real world showing the effective measurement height (z_f) and UKV grid box dimensions (1.5 km), and (b) model world's (UKV) effective surface z_{ES} and atmospheric levels. In (a) a hypothetical tower with EC sensor is shown in the centre of the LAS path. (c) The cross-wind and along-wind 60% SA weights are shown for EC (blue) and LAS (red) sensors for one wind direction (θ, perpendicular to the LAS path) and stability (unstable). [Colour figure can be viewed at wileyonlinelibrary.com]

In this study, we consider from both the modelling and observation perspective how to evaluate NWP models (Section 2) using LAS observations over highly complex surfaces like cities. Despite its advantages, LAS has only been used to evaluate mesoscale NWP models over non-urban surfaces (e.g., Beyrich et al., 2002; Steeneveld et al., 2011). We introduce a methodology to address the following questions:

1. How can consistency between LAS observations and modelled Q_H be established with respect to the spatial and temporal scales involved?
2. How sensitive are evaluation outcomes to LAS averaging times and spatial overlap between LAS SAs and kilometre-scale model grid lengths?

We demonstrate our approach using the Met Office's operational NWP model for the United Kingdom, which is a regional configuration of the Unified Model (MetUM) known as the UKV (Tang et al., 2013; Section 4.2.2), for two intensive observation period (IOP) days with LAS observations in central London (Section 4.1). From this analysis (Section 4), we recommend how the approach can be applied in future longer-term model evaluation studies.

2 | CONSISTENCY BETWEEN OBSERVATIONS AND MODEL OUTPUT

When comparing models and observations, it is important to take account of the effective horizontal (Section 2.1), vertical (Section 2.2), and temporal (Section 2.3) resolution of both to ensure consistency. After this general discussion, it is applied to a case study (Section 4).

2.1 | Horizontal consistency

A sensor's SA, footprint, or field of view indicates the probable area of influence impacting a measured variable, but varies with sensor type (e.g., sonic anemometer vs LAS for Q_H). This area can be modelled, allowing the characteristics of the area identified (e.g., influencing radiation and sensible heat fluxes, Vesala et al., 2008) to be used to assign appropriate model parameters (Grimmond & Oke, 1991). Turbulent flux SA shape, location, and size vary with sensor height (e.g., Schmid & Lloyd, 1999), wind direction, atmospheric stability and surface roughness characteristics. For example, SAs may extend further upwind of a measurement point in stable *cf.*

unstable conditions (e.g., Leclerc & Foken, 2014). As the real world is not perfectly homogeneous, LAS-path effective measurement height (z_f) can change with meteorological conditions because of changes in the SA characteristics (Section 3.1.1).

LAS source area extent and shape are a function of path orientation, length, and path weighting (e.g., Evans et al., 2012; Meijninger et al., 2005; Ward et al., 2014). SA_{LAS} with O (1–10 km^2) spatial extents are typically much larger than the O (0.10–1 km^2) EC SAs (fig. 1 in Crawford et al., 2017; *cf.* fig. 1. in Ward, 2017), reflecting the line vs point measurement types (Figure 1). Such differences are most evident when the path is perpendicular to the prevailing wind direction (Meijninger et al., 2002). For example, the width of the unstable (stability parameter $\zeta < -0.03$, see Section 3.1) SA_{LAS} is 3.6 km but only 0.57 km for SA_{EC} (Figure 1c) with both sensors at the same height and 60% probability (Figure 1a). Thus, the SA_{LAS} can be of the same order as typical regional NWP model resolutions. With the UKV grid length of 1.5 km (Tang et al., 2013), the SA_{LAS} lateral extent covers almost three grid boxes, whereas the SA_{EC} lateral extent is about half a grid box (Figure 1c). The larger SA_{LAS} therefore reduces the scale gap existing between NWP and typical flux observations (Kleissl et al., 2009).

2.2 | Vertical consistency

The lower boundary of the inertial sublayer (ISL, Figure 1), or blending height (z_r), is critical to both the observed and modelled fluxes. The depth of the urban canopy layer (UCL, Figure 1a) is controlled by the height of the roughness elements (e.g., buildings or trees), with individual roughness elements modifying flow and exchange processes in the roughness sublayer (RSL). Above these, in the ISL (or constant flux layer), the impact of individual roughness elements is blended to give a local scale or neighbourhood response (Grimmond & Ward, 2021), and little height variation of the turbulent momentum and scalar fluxes (Katul et al., 1995; Moene & Schüttemeyer, 2008).

For both the model and observations, a logarithmic wind profile and Monin–Obukhov similarity theory (MOST) are assumed, and therefore the comparison height must be within the ISL. To derive the large-aperture scintillometry sensible heat fluxes ($Q_{H,LAS}$), the effective measurement height (z_f) needs to be in the ISL (Figure 1a).

NWP model levels typically have increasing spacing with height above ground level (Tang et al., 2013), with each grid box having one topographic altitude. In the real world, areas with taller roughness elements have a greater mean height, and the momentum sink is higher above ground level if the displacement height (z_d) increases (e.g., Oke et al., 2017). The momentum roughness length of the surface (z_0) may also increase, but depends on the spacing of the roughness elements. Most NWP models have a land surface scheme providing parametrised RSL processes to the lowest atmospheric model level, often assumed to be at or above z_r (i.e., in the ISL or constant flux layer). This assumption (Masson, 2006, see their fig. 1) is made in the case of the UKV (Figure 1), with the follow-on assumption that MOST holds, allowing screen-level air temperature and 10 m wind speed to be diagnosed. The former is typically calculated for the nominal height of the first model level (Masson, 2006). With short roughness elements (e.g., short green grass) this presents no additional problems.

2.3 | Temporal consistency

$Q_{H,LAS}$ are derived from much larger spatial domains than eddy covariance derived Q_H ($Q_{H,EC}$) (Beyrich et al., 2021). For $Q_{H,LAS}$, it is possible to have much shorter averaging periods (e.g., 1 min *cf.* ≥30 min for $Q_{H,EC}$), as LAS does not need to sample all relevant energy-carrying eddies (*cf.* $Q_{H,EC}$) because the sensor is sensitive to the eddy size corresponding to the aperture diameter/the first Fresnel zone (typically 0.1–0.3 m). Thus, this eddy size is much smaller than the LAS path length (typically 0.5–5 km), and therefore a statistically stable flux can be derived in a shorter timeframe (De Bruin et al., 1993). The SA_{LAS} will change at the temporal frequency of the flux calculation period, as stability and other meteorological conditions vary.

While 1-min $Q_{H,LAS}$ can be compared to instantaneous NWP-Q_H, it is qualitatively not appropriate because at the UKV's relatively coarse spatial resolution, it does not resolve explicitly smaller-scale atmospheric motions that cause the variability seen in 1-min $Q_{H,LAS}$. This is explored in our results (Section 6.2.3).

3 | METHODS

The methodology employed to process LAS data involves deriving both the SA_{LAS} and the turbulent sensible heat flux ($Q_{H,LAS}$). A case study is used to demonstrate the proposed evaluation method. Python software developed for the LAS processing and model evaluation can be accessed via the links provided in the Supporting information, Table SM.1.

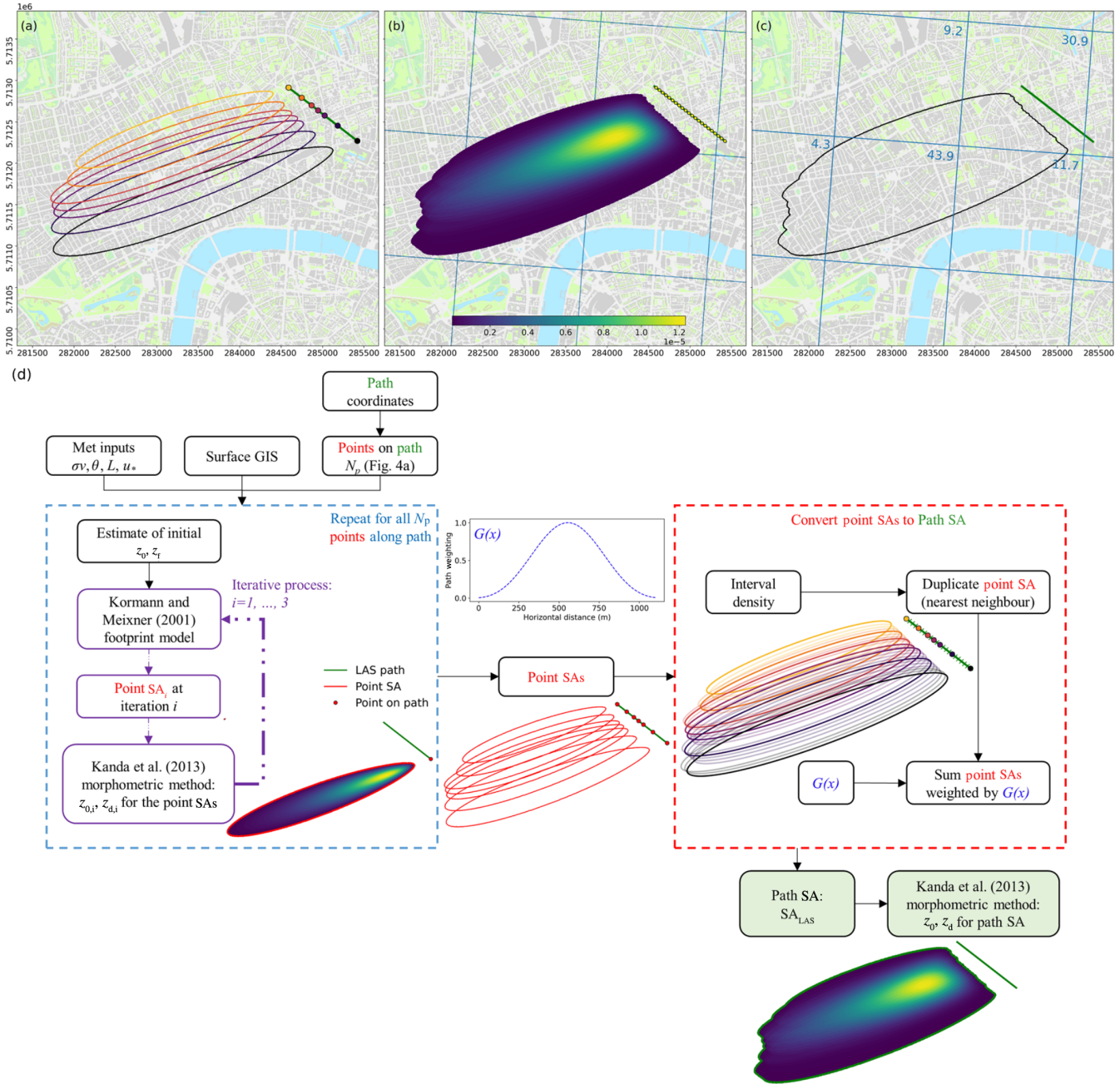

FIGURE 2 Example of large-aperture scintillometry (LAS) source area (SA) calculation in London under unstable conditions (1200 UTC on 21 May 2016). Workflow of (a) and (b) shown in (d). (a) Seven ($N_p = 7$) unique eddy covariance (EC) SA are calculated along the LAS path (green) at the dots indicated, (b) these are replicated outwards at 50 m (yellow dots) along the LAS path (green), with the SA_{LAS} weight (colour bar) per pixel of resolution of 4 m^2 underlain by the UKV 1.5 km grid boxes (blue quadrilateral); and (c) a 60% SA_{LAS} extent (black) with weights within the overlain NWP grid box. (d) Flowchart of path SA calculation. Notation given in Table 1. Surface GIS includes building/terrain/vegetation heights and landcover fractions. [Colour figure can be viewed at wileyonlinelibrary.com]

3.1 | Large aperture scintillometry observations

3.1.1 | LAS source area calculation

The general steps, improving on earlier work, to derive SA_{LAS} over cities are presented in Figure 2d.

The proposal by Meijninger et al. (2002) to adapt EC flux footprint models for LAS paths by combining single-point measurement SAs with the LAS path-weighting function, has been used with different EC footprint models (e.g., Evans et al., 2012; Ward et al., 2014; Ward, Evans, & Grimmond, 2015). Like Crawford et al. (2017), we use the Kormann and Meixner (2001)

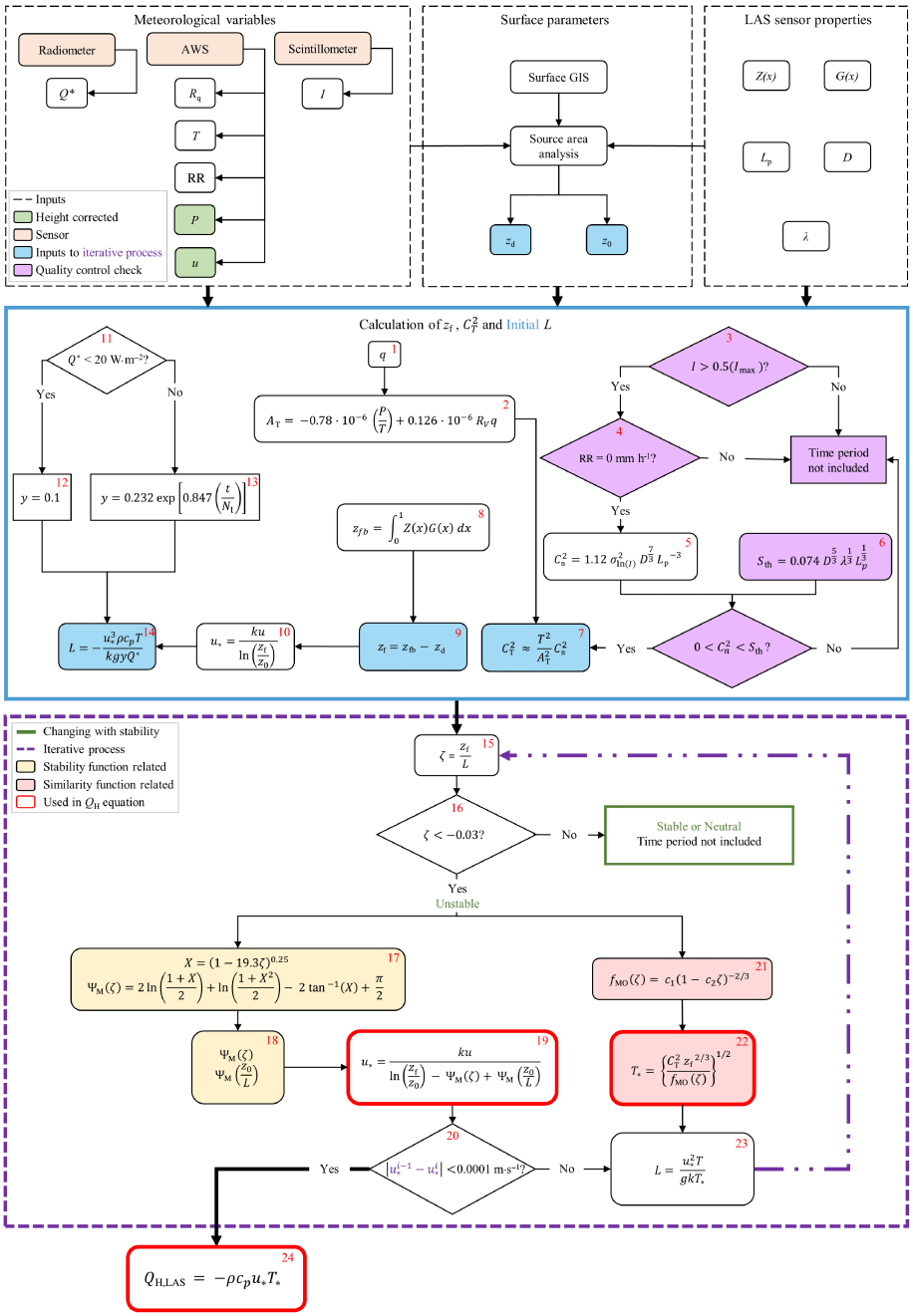

FIGURE 3 Flowchart of $Q_{H,LAS}$ calculation with steps numbered (red). Notation and references given in Table 1. More details on surface parameters given in Figure 2d. Software is available, with location listed in Table SM.1. [Colour figure can be viewed at wileyonlinelibrary.com]

footprint model. This SA_{EC} model calculates the probable SA for an EC sensor at a point (x, y, z) using meteorological data for a time interval (Figure 3) and surface parameters (roughness length z_0, and displacement height z_d). Three heights are needed (Figure 1): (a) the height above ground of the sensor (z), (b) z_d and (c) z_0. Uncertainty in calculating the two aerodynamic roughness parameters is one constraint in determining SA_{LAS}, as representative values are difficult to determine in complex terrain like cities given large variations of roughness element height (e.g., Ward, 2017).

To determine the SA for a LAS path, we first calculate multiple SA_{EC} for different points along the path (Figure 2a,d). The location and number of unique SA_{EC} source points (N_p) needs to reflect the LAS path weighting (Figure 2d), with fewer N_p needed as the weighting function approaches zero at the path ends. This centre-weighted sensitivity is approximated by the scintillometer path weighting function $G(x)$ (Evans & de Bruin, 2011):

$$G(x) = 2.163\left(\frac{2J_1(2.283\pi(x-0.5))}{2.283\pi(x-0.5)}\right)^2 \quad (3)$$

where J_1 is a Bessel function of the first kind and x is a given distance along the LAS path.

Crawford et al. (2017) assume z_0 and z_d do not vary along the LAS path. However, Kent et al. (2017) demonstrate that the morphometric method chosen to calculate z_0 and z_d impacts modelled SA_{EC}. They show if the z_0 and z_d methods account for roughness element height variability, for example, Kanda et al.'s (2013) method, the SA_{EC} modelled are a third (or smaller) of the SA_{EC} size obtained if only the mean roughness element height is considered (e.g., Grimmond & Oke, 1999; Macdonald et al., 1998; rule-of-thumb method). This is because of the lower effective measurement height ($z_f = z - z_d$, Figure 3 – eq. F.9) linked to the larger z_d typically found in areas of strong building height variability.

When calculating the SA_{EC} along the LAS path, z_0 and z_d are determined iteratively (Figure 2d), following Kent et al. (2017), who show that appropriate values are obtained after three iterations using the Kanda et al. (2013) method, independent of initial values selected. Results of z_0 and z_d along with geographic information system (GIS) surface information are used within a SA analysis to characterise the surface influencing the observations.

Here, for point SA analysis, initial values of $z_0 = 1.5$ m and $z_f = 80$ m are used. Each SA_{EC} is calculated for a cumulative weight of interest (here: 60%). Given a high spatial resolution of surface cover data (e.g., 4 m^2), each pixel area contributes a small amount. The minimum weight is 1.9×10^{-10}. Larger cumulative weights increase the computational cost (Crawford et al., 2017; Kent et al., 2017).

Each SA_{EC} determined for the N_p points are replicated in both directions along the LAS path at regular intervals using a nearest-neighbour approach (Figure 2b,d). This is to increase the density of surface information along the path for the final weighting. In Figure 2b, an interval density of 50 m is used for a path length of 1112 m. Finally, the SA_{LAS} (Figure 2b) is obtained from the sum of individual EC SAs along the path weighted by $G(x)$. The aerodynamic parameters used to obtain the $Q_{H,LAS}$ (Section 3.1.2; Figure 3, eq. F.24) and the SA_{LAS} are calculated using the Kanda et al. (2013) method.

However, whether or not this cumulative EC flux footprint approach provides a reasonable SA estimate for the LAS-derived structure parameter (i.e., the primary LAS observation, Table 1/Figure 3) for the derived sensible heat flux is unknown. Ward (2017) suggest uncertainties that arise from this conceptual discrepancy are likely smaller compared to errors from the application of the SA model over a heterogeneous surface. Future research on structure parameter footprints could provide valuable insights into this issue.

3.1.2 | LAS sensible heat fluxes derivation

Beyond the LAS beam intensity (I), the meteorological variables required to obtain $Q_{H,LAS}$ are relative humidity (R_q), air temperature (T), air pressure (P), wind speed (u), wind direction (θ) and net all-wave radiation (Q_*) (Figure 3). Initial quality control steps remove LAS data if rain (>0 mm h^{-1}, eq. F.4) or fog (LAS beam intensity $I < 0.5 I_{max}$, eq. F.3, where I_{max} is the daily maximum beam intensity) occur. Measurements are also excluded if saturation occurs and scintillations exceed a given threshold (S_{th}) in the refractive index structure parameter C_n^2 (Kohsiek et al., 2006). S_{th} is defined in Figure 3, eq. F.6 (Kleissl et al., 2010).

Additionally, the flux calculation requires surface characteristics derived from the SA analysis (Section 3.1.1), notably the LAS-path effective measurement height z_f (Figure 3, eq. F.9). Along with z_f, the initial estimates of the structure parameter of temperature (C_T^2, eq. F.7), and stability (eq. F.14) are inputs (Figure 3, blue) to an iterative process (Figure 3, purple, eqs F.15–23) which updates the stability parameter until the friction velocity (u_*) meets the convergence criterion (difference $<10^{-4}$ m s^{-1}, eq. F.19). Once this threshold is met, $Q_{H,LAS}$ is calculated (eq. F.24). The iterative loop continues only if the first iteration yields $\zeta < -0.03$ (eq. F.15), indicating unstable stratification. As during stable conditions, with weak turbulence and a shallow boundary layer, it is more challenging to guarantee sufficient turbulence reaches the sensor path requiring additional quality controls of flux validity (Beyrich et al., 2021). Furthermore, even for slightly stable or neutral conditions, the sign of $Q_{H,LAS}$ cannot be determined unambiguously (Ward, 2017).

It is assumed (eq. F.22) that the LAS-path z_f is within the ISL (Figure 1a), so the logarithmic wind profile (eqs F.10 and F.19) and MOST are applicable. Other constraints exist with determining LAS fluxes in complex environments (Ward, 2017). Under very stable conditions ($\zeta > 0.03$), MOST assumptions no longer apply, and LAS sensors will not measure sufficient turbulence at z_f for a flux to be determined (Hartogensis et al., 2002). This becomes more problematic with rougher surfaces (De Bruin et al., 1999; Katul et al., 2006; Moene & Schüttemeyer, 2008). As MOST similarity functions are used to relate the structure parameter of temperature C_T^2 to the sensible heat flux, these introduce additional uncertainties (e.g., Kooijmans & Hartogensis, 2016; Maronga, 2014; Ward, Evans, & Grimmond, 2015), linked to their empirically derived coefficients (Figure 3, eq. F.21) and approximations.

TABLE 1 Calculation of $Q_{H,LAS}$ uses Figure 3 equations (red numbers, here column F).

F	Symbol	Description	Units	Note, more details	Reference
1	q	Absolute humidity	kg m^{-3}	App. A	McIlveen (2010); Vaisala (2013)
	R_q	Relative humidity	%	To calculate q, App. A	
2	A_T	Structure parameter coefficient for temperature	–		Andreas (1988)
	P	Air pressure	Pa	Height correction, App. A	
	T	Air temperature	K		
	R_v	Gas constant for water vapour	J K^{-1} kg^{-1}	= 461.5	
3	I	Beam intensity	–	I_{max} daily maximum I	Crawford et al. (2017)
4	RR	Rain rate	mm h^{-1}		Crawford et al. (2017); Ward et al. (2014)
5	C_n^2	Refractive index structure parameter	m$^{-2/3}$		Wang et al. (1978)
	D	Instrument's Fresnel lens diameter	m		
	L_p	Path length between transmitter and receiver	m		
6	S_{th}	Saturation threshold	m$^{-2/3}$		Kohsiek et al. (2006), Kleissl et al. (2010)
	λ	Wavelength	m		
7	C_T^2	Structure parameter of temperature	K^2 m$^{-2/3}$	Simplified, App. B	Wesely (1976), Hill et al. (1980)
8	z_{fb}	Effective beam height	m		Hartogensis et al. (2003)
	$Z(x)$	Beam height along path	m		
	$G(x)$	Path weighting function	–	Equation (3)	
9	z_f	Effective measurement height	m		Hartogensis et al. (2003)
	z_d	Displacement height	m		
10	u_*	Friction velocity	m s^{-1}	Initial: neutral stability	Stull (1988)
	z_0	Roughness length	m		
	u	Horizontal wind speed	m s^{-1}	Height correction, App. A	
	k	von Kármán's constant	–	= 0.4	
11	Q^*	Net all-wave radiation	W m^{-2}		Grimmond and Cleugh (1994)
12	y	Q^* scaling factor ($Q_H \approx yQ^*$) to obtain first Q_H estimate for L (F.14)	–	F.12 if $Q^* < 20$ W m^{-2}	
13				F.13 if $Q^* \geq 20$ W m^{-2}	
	t	Hour number after $Q^* > 20$ W m^{-2}	h	First hour that $Q^* > 20$ W m^{-2}: $t = 1$ Second hour: $t = 2$, etc.	
	N_t	Number of hours in day $Q^* > 20$ W m^{-2}	–		
14	L	Obukhov length	m	Initially: $Q_H \approx yQ^*$	
15	ζ	Stability parameter	–		Foken (2008)
16		Stability classification		Unstable: $\zeta < -0.03$	Ward (2013, their section 8.6)
17	$\Psi_M(\zeta)$	Integrated form of the stability function for momentum		Unstable	Van Ulden and Holtslag (1985)
18					
19	u_*	Friction velocity	m s^{-1}		Stull (1988)
20	u_*^{i-1}	iteration (i); prior ($i-1$)			
21	$f_{MO}(\zeta)$	Similarity function	–	Unstable	Dyer (1974); Foken (2008); Wyngaard et al. (1971)
	c_1	Similarity function constant	–	= 4.9	Andreas (1988)
	c_2		–	= 6.1	
22	T_*	Temperature scaling variable	K		Crawford et al. (2017)
23	L	Obukhov length	m		Ward (2013)
24	$Q_{H,LAS}$	Sensible heat flux	W m^{-2}		Ward (2013)

Note: Same empirical coefficients (e.g., eq. F.21) as Crawford et al. (2017). Same definition as Ward (2013) (their section 8.6) for unstable ($\zeta < -0.03$) conditions.

3.2 | Numerical weather prediction model

To demonstrate the evaluation approach, we use UKV output. This is the Met Office operational regional forecast model for the United Kingdom, with 1.5 km horizontal grid length, that is one-way nested within the 17 km global model (Tang et al., 2013). The UKV uses the land surface model JULES (Joint UK Land Environment Simulator, Best et al., 2011; Clark et al., 2011) with the urban scheme MORUSES (Porson et al., 2010a, 2010b).

JULES uses a tiling approach to represent the variability of the land cover at the subgrid-scale (SGS) with five vegetation tiles and five non-vegetated tiles in the UKV configuration, including two tiles to calculate the fluxes from impervious surfaces using MORUSES. This assumes that built parts of cities can be represented by an infinitely long street canyon, consisting of a two-dimensional canyon tile and a roof tile (Harman et al. 2004a; Porson et al., 2010b). The effective parameters (i.e., heat capacity, albedo, and emissivity) are calculated based on canyon morphology (i.e., building height $\overline{z_b}$, canyon width, height to width ratio) and using prescribed material properties for clay roofs, brick walls and asphalt road facets (Bohnenstengel et al., 2011; Table SM.2). The resistance network varies with flow regimes defined by morphology (Harman et al. 2004b). Q_H is calculated separately for the canyon and roof, and the other land cover tile fractions (Figure 4b) to give the aggregate value for each grid box, $Q_{H,UKV}$ (Best et al., 2011). It is updated at the end of each model time step using implicit coupling between JULES and the atmospheric model.

The UKV uses a hybrid height-based vertical coordinate system; this is terrain-following at the surface, and flat at the top. Distances between model levels vary (smoothly) between a quadratically stretched Charney–Philips staggered grid (Tang et al., 2013).

Aerodynamic roughness parameters (z_0 and z_d) used in MORUSES are derived using Macdonald et al.'s (1998) morphometric method. However, z_d is only used to calculate z_0 and not currently applied in the blending scheme for the coupling to the atmospheric model. Conceptually, this implies that buildings are displaced into the ground by z_d with an effective surface (z_{ES}, Figure 1) at z_0. UKV–MORUSES wind speed is assumed to be $0\,m\,s^{-1}$ at a height of $z_d + z_0$. Thus, the first model level of the UKV is within the ISL (Figure 1, Section 2.2). z_{ES} is the height at which surface fluxes are calculated and from where the UKV model levels are measured, with the first model level at 2.5 m above z_{ES} plus the local orography (Boutle et al., 2021). This first level (i.e., $z_{ES} + 2.5$ m) is acceptable in areas of low roughness (e.g., short grass), but is conceptually challenging in areas with large roughness elements (e.g., cities or forests).

Anthropogenic heat flux, Q_F, is prescribed in MORUSES with fixed monthly values that scale with the total impervious land cover fraction (e.g., Hertwig et al., 2020), rather than varying in response to actual daily temperature, day of week, or time of day (e.g., Capel-Timms et al., 2020). Latent heat flux is modelled for vegetated surfaces with the appropriate JULES tile, but leaf area index may not be correct in the urban area (Hertwig et al., 2020). For the two built tiles, latent heat flux is relevant only immediately after precipitation events (Hertwig et al., 2020). The grid box value is again weighted by surface cover fractions present (Best et al., 2011).

4 | EVALUATION PROCEDURE

The procedure for evaluating modelled $Q_{H,UKV}$ when $Q_{H,LAS}$ observations are available is demonstrated in central London (Figure 4) for two days. The two IOP days have different incoming short-wave radiation, wind direction, wind speed and stability (Figure 5). IOP-1: 2 May 2016 is cloudy with the wind direction perpendicular to the LAS path (~210°), and IOP-2: 5 May 2016 is clear with the wind direction along the path (~130°).

4.1 | Site description

The study area in central London is densely built (Figure 4b). The surface characteristics within the SA_{LAS} are derived from ~4-m resolution raster digital surface model (DSM) data (Lindberg & Grimmond, 2011; Figure 4a). The UKV land cover at 1.5 km resolution is derived from an older 25-m resolution dataset (CEH, 1990). While there is variability between land cover fractions across the UKV grid boxes (Figure 4b), it is relatively little, and the ratio of the canyon tile to roof tile is fixed by an empirical relation (Bohnenstengel et al., 2011). Both datasets differ from the 'reality' (e.g., Figure 4c) that the LAS sensors are exposed to (e.g., buildings added/removed with time, resolution of acquired data; Morrison et al., 2021).

More vegetated (Figure 4a) UKV grid boxes, with major parks (e.g., St. James' Park; grid box H), have higher C3 grass proportions (Figure 4b), whereas the UKV grid boxes including the River Thames (Figure 4a; H, E, F) have a land cover proportion classified as inland water. The largest inland water proportion is 14% (Figure 4a; E).

UKV grid box vegetation proportions (details in Table SM.4) vary from 3% (D, F) to 14% (H). This differs from the observation reference dataset (hereafter VL) (Lindberg & Grimmond, 2011) grid box values (Table SM.5); grid box D has a vegetation proportion of

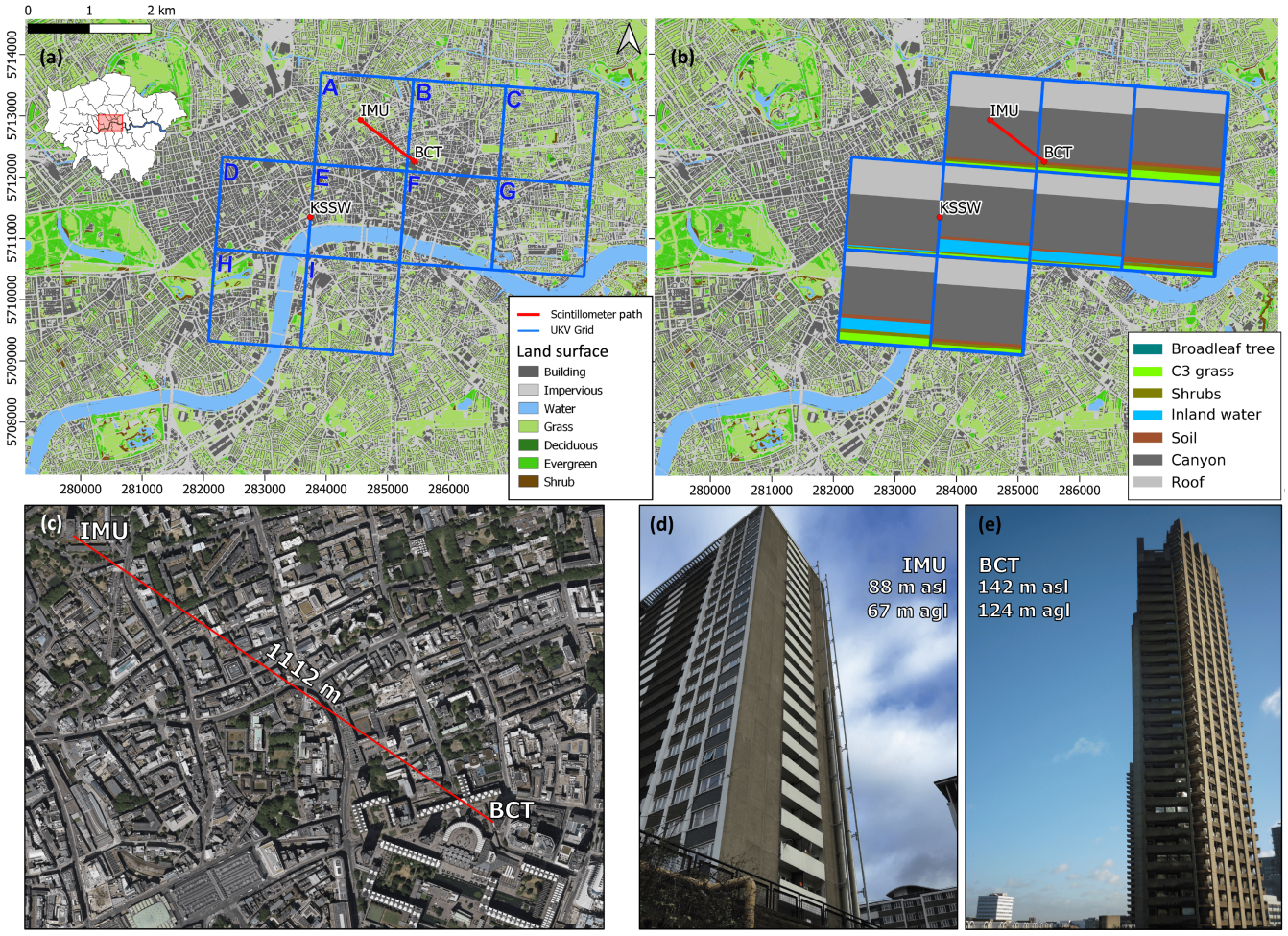

FIGURE 4 Study area with large-aperture scintillometry (LAS) path (red line), (a) land cover (based on Lindberg & Grimmond, 2011) shown in UTM zone 31 N (m) coordinates, and UKV grid boxes (blue, identified in text by labels A–I) and (inset) location in London (red rectangle); (b) UKV land cover proportions (Table SM.4); (c) aerial image (Bing Maps, 2021) of LAS path (red) region; and location of LAS (d) receiver (IMU) and transmitter (BCT). [Colour figure can be viewed at wileyonlinelibrary.com]

14%, F = 11%, and H = 35%. The building (roof) proportion in the UKV varies from 17% (H, VL:13%) to 39% (D, VL:31%) and canyon from 53% (H, VL:30%) to 60% (G, VL:46%) (Figure 4b, Tables SM.4 and SM.5).

4.2 | Data description

4.2.1 | Instruments

A single-beam Kipp and Zonen MKII LAS (850 nm wavelength) sensor is used, with the transmitter (Tr, Figure SM.1a) located 1112 m (Figure 6) from the receiver (Rx, Figure SM.1b,c). The Tr mounted 142 m above sea level (a.s.l.) at the BCT site (Figure 4e) in the City of London, has a Davis Vantage Pro Plus sensor measuring other meteorological variables at the same height (Figure SM.1a). The LAS receiver is mounted 88 m a.s.l. at the IMU site (Figure 4d) in the Borough of Islington. Figure 6 shows the height differences, together with the transect, of the building heights directly below the LAS path. The tallest buildings are approximately half the height of the building where the receiver is mounted.

A radiometer (CNR4, Kipp & Zonen) measures the four constituents of Q^* (i.e., long-wave and short-wave incoming and outgoing radiation, sampled at 0.2 Hz) at 61 m a.s.l. at the KSSW site (Figure 4). Kotthaus & Grimmond (2014a, 2014b), who use the same sensor at the same location as this study, describe the radiometer field of view of the upwelling fluxes in detail. Quality controls for $Q_{H,LAS}$ include data removal during rain, fog, and saturation occurrences (Section 3.1.2, Figure 3). The processing steps for the meteorological data needed to derive $Q_{H,LAS}$ are discussed in Appendix A.

After sampling the LAS signal intensity (I) at 500 Hz, the mean 10-Hz logarithms, ln(I), are stored. The variance

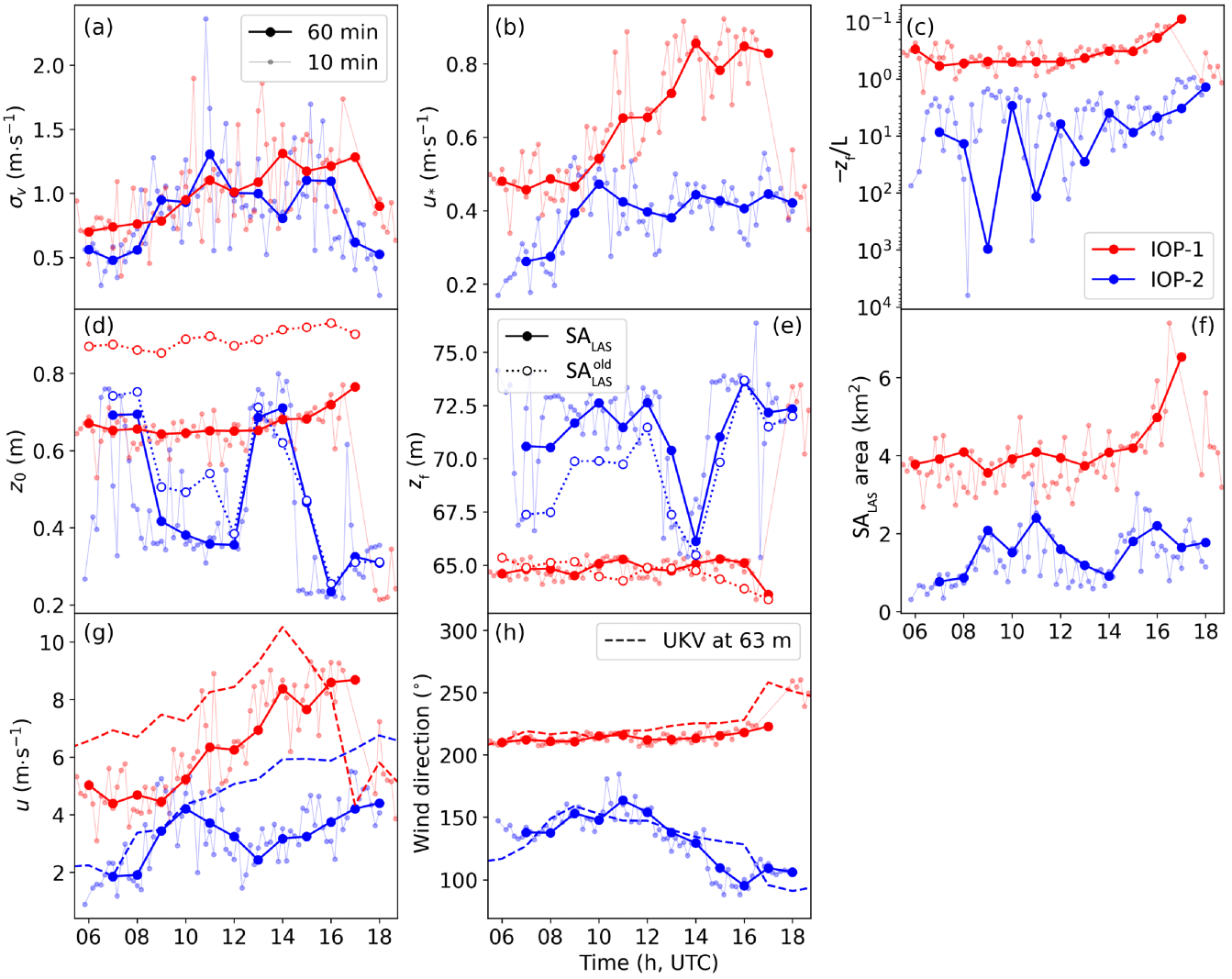

FIGURE 5 Variables during the unstable period ($z_f/L < -0.03$) on two IOP days (colour) determined for 10-min (small symbols) and 60-min periods (large symbols): (a) lateral velocity standard deviation (σ_v), (b) friction velocity (u_*), (c) stability ($-z_f/L$), (d) roughness length (z_0), (e) effective measurement height (z_f), (f) overall extent of the 60% source area (SA_{LAS}) weight, (g) wind speed (u), and (h) wind direction. UKV model prognostic wind direction and wind speed (dashed, hourly samples) at the model level closest to the observation height ($z_f - z_0$) included in (g,h). Roughness parameters from two different SA_{LAS} calculation methods are presented (d,e); the Crawford et al., 2017 method uses one SA_{EC} (SA^{old}_{LAS}), whilst the proposed method uses seven different SA_{EC} (SA_{LAS}). [Colour figure can be viewed at wileyonlinelibrary.com]

of ln(I) is calculated using these 10-Hz data for the averaging period of interest (e.g., 5 min) prior to calculating C_n^2 (step 5, Figure 3) (Crawford et al., 2017; Ward, 2013).

For each IOP day, we calculate $Q_{H,LAS}$ (and compare to the UKV) during unstable stratification (Section 3.1.2). Periods analysed for each IOP day are the first to last 10-min period determined to be unstable.

$Q_{H,LAS}$ is calculated for 1-, 5-, 10-, and 60-min time-ending periods. Given the SA_{LAS} calculations' computational costs (>300 s per SA_{LAS}), these are calculated every 10 and 60 min to derive z_f. For $Q_{H,LAS}$ averaging periods of 10 min or less, the appropriate 10-min z_f is used, whereas for the 60-min $Q_{H,LAS}$ the 60-min SA_{LAS} z_f are used.

4.2.2 | UKV model output

The UKV data for the IOP days are from the operational runs, initialised at 2100 UTC, with model output analysed from 0000 UTC for 24 h for the period when observations are available. Therefore, the first three hours are treated as spin-up (e.g., Warren et al., 2018). The operational UKV's diagnostics output is on the hour. The averaging time for UKV-modelled Q_H is the same as the model timestep (1 min). For $K_\downarrow$ at the surface, a 15-min average (time starting) is output.

4.2.3 | Observation uncertainty

Uncertainties are introduced from multiple sources during the process of deriving $Q_{H,LAS}$ (Figure 3), such as LAS

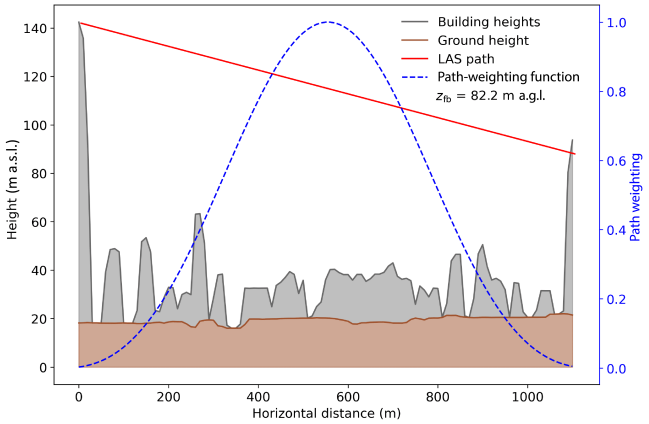

FIGURE 6 LAS path-weighting function (blue dashed, Equation 3) with building (grey) and ground (brown) heights along a LAS path (red) with the transmitter and receiver. For the London case the effective beam height (z_{fb}, eq. F.8, Figure 3) is 82.2 m above ground level (a.g.l.). Note, z_{fb} differs from the effective measurement height (eq. F.9, Figure 3) and does not account for displacement height. Digital surface model (Lindberg & Grimmond, 2011) building heights analysed at 10-m resolution. [Colour figure can be viewed at wileyonlinelibrary.com]

observation errors, SA attribution, instrument uncertainty and height, path saturation, application of similarity theory, and other meteorological measurements. These uncertainties often interact, resulting in individual uncertainty sensitivity testing being challenging. Ward (2017) suggests that a direct analysis of LAS-measured structure parameters, with fewer processing steps than $Q_{H,LAS}$, is advantageous for attributing uncertainty sources. The uncertainties and assumptions associated with steps to calculate $Q_{H,LAS}$ in Figure 3 (eqs F.1–24) include:

a *LAS measurement error, sensor properties and meteorological variables*: If LAS beam saturation occurs, the turbulence is insufficient for the intensity fluctuations to still be proportional to C_n^2 (Figure 3, eq. F.5), affecting $Q_{H,LAS}$ (Kleissl et al., 2010). This is mitigated by applying a saturation filter (Figure 3, eqs F.3–6) but uncertainty increases towards the saturation threshold (Figure 3, eq. F.3). Reducing path length, increasing path height, and/or using a larger aperture sensor helps to mitigate this (Ward, 2017). Uncertainty associated with each sensor type (radiometer, automatic weather station/AWS, LAS pair), and their corresponding variables (including processing steps and quality control), will introduce uncertainty in the derived $Q_{H,LAS}$. Sensitivity analysis of the impact of meteorological variables (pressure, air temperature, wind speed) to the derivation of $Q_{H,LAS}$ by Crawford et al. (2017, their Figure 6), found the largest impact when changing wind speed by ± 1 m s^{-1} (~3.5% impact on $Q_{H,LAS}$). If sensor SA characteristics differ (e.g., radiometer vs LAS path) additional uncertainty is introduced. As radiometer SAs are much smaller than SA_{LAS}, using multiple radiometers may ensure they are both within and representative of each SA_{LAS}.

b *Surface parameters*: Uncertainty is introduced to $Q_{H,LAS}$ through the methods and assumptions made when calculating the aerodynamic roughness parameters from the SA_{LAS}. First, the SA model used (Kormann & Meixner, 2001) provides a probabilistic SA_{EC} for a selected extent (60% in this case) and its application along the LAS path (SA_{LAS}, Figure 2) contributes to the uncertainties in effective beam height and roughness parameters. By accounting for variations of these parameters along the beam (Section 3.1.1, Figure 2) we assume we mitigate some of this uncertainty (*cf.* e.g., Crawford et al., 2017). Second, it is unknown if the structure parameter (Section 3.1.1) SA is appropriately characterised by the SA_{LAS}.

c *Structure parameters*: The calculation of LAS structure parameters (Figure 3, eqs F.5, F.7) assumes the coefficient for temperature (A_T, Figure 3, eq. F.2) is significantly larger than for humidity (A_q see Appendix B, impacting the structure parameters and therefore $Q_{H,LAS}$). As A_q is expected to be larger during rain periods when the surface is wet, this is mitigated by removing rain events (Figure 3, eq. F.4) and following rain (e.g., ≥ 3 h after rain, Crawford et al., 2017).

d *Similarity functions*: The similarity functions (Figure 3, eq. F.21) used to relate the structure parameters to Q_H have large uncertainty (e.g., Kooijmans & Hartogensis, 2016, Section 3.1.2; Maronga, 2014; Ward, Evans, & Grimmond, 2015). Frequently, Andreas' (1988) (Figure 3, eqs F.15–24) and De Bruin et al.'s (1993) methods are used, which can create differences of up to 15% in $Q_{H,LAS}$ (Beyrich et al. 2012; Crawford et al., 2017; Lagouarde et al., 2006; Meijninger et al., 2005; Ward et al., 2014). The uncertainty arises from the coefficients chosen (Figure 3, eq. F.21, $c_{1,2}$), with the most appropriate choice for urban areas still being a research topic (Crawford et al., 2017).

e *Applicability of Monin–Obukhov Similarity Theory over complex terrain*: The similarity functions (i.e., point d. above) assume MOST is applicable and therefore that the observations are at sufficient height for the surface effects to be blended (homogeneous) and for a duration that ensures stationarity (Ward, 2017). Thus, it is essential to ensure the LAS path is in the inertial sublayer (ISL, Figure 1), where the influence of individual roughness elements is assumed to be blended (e.g., momentum, Foken, 2008). However, it remains uncertain if the blending height concept is also applicable to structure parameters (Beyrich et al., 2012). MOST

assumptions are harder to satisfy during stable than unstable conditions (Ward, 2017), another reason and benefit for our focus on unstable conditions (Figure 3, eq. F.16).

4.3 | Comparison approach and metrics

$Q_{H,UKV}$ is compared to $Q_{H,LAS}$ at two model levels: (1) surface/first model level (Figure 1) with the horizontal spatial extent informed by the SALAS analysis (Section 3.1.1 and (2) the level closest to the time-varying LAS effective height for each observation to assess vertical consistency (Section 2.2) for the grid box overlaying the centre of the path. For (1) the SA_{LAS} is used to identify the contained model grid boxes for a timestep, allowing the weighting for the mean surface $Q_{H,UKV}$ (Figure 2c). For comparison with the model (Section 6), the LAS flux for a particular averaging period is selected at the end of the hour. To assess the impact of temporal resolution (Section 2.3) $Q_{H,UKV}$ is compared to $Q_{H,LAS}$ determined for different periods (1, 5, 10, and 60 min).

The UKV grid box flux is the land cover fraction weighted mean for each JULES surface tile, representing the blended flux at the model's first atmospheric level. This surface flux is by definition in the ISL (Mason, 1988) and assumed to be at a height of $z_0 + z_d + 2.5$ m above ground level in the UKV (Figure 1). Use of both the LAS and the model rely on assumptions (e.g., MOST) that are only valid in the ISL, hence both $Q_{H,LAS}$ and $Q_{H,UKV}$ are assumed to be in the ISL. We need to identify what the model output represents (constrained by the model vertical levels) and what the effective height of the observations is. We do not adjust the height of the observed fluxes or the observations for height.

From the model perspective, these are two heights that may be appropriate, both within the ISL. Under unstable conditions during the middle of the day, the urban boundary layer should be sufficiently deep that all model levels below the effective LAS measurement height are within the surface layer (i.e., from 0 m to ~10% of the boundary-layer depth, Oke et al., 2017). Given there is an expectation that within the ISL (or constant flux layer) the fluxes vary by less than 10% vertically (Oke et al., 2017), we also assess that (Section 6.2.2).

Evaluation metrics (Table SM.7) are used to quantify the differences/similarities between $Q_{H,LAS}$ and $Q_{H,UKV}$. These include the absolute error (AE) and the bias error (BE) for each instance. The BE sign indicates if $Q_{H,UKV}$ over- or underpredicts the observed flux. Additionally, the means (MAE, MBE) are given. All have the units of the variable, with a perfect result being 0 (e.g., 0 W m^{-2} for Q_H). The interquartile range (IQR) is calculated for some variables throughout Section 6. All times referred to are UTC.

5 | OBSERVATION RESULTS

5.1 | Source area characteristics

To assess the horizontal consistency between $Q_{H,LAS}$ and $Q_{H,UKV}$, we first address the characteristics of the SA_{LAS} on both IOP days. As noted (Section 3.1.1), SA_{LAS} vary with time, stability (z_f/L) and wind direction, and are determined as part of the $Q_{H,LAS}$ calculations (Section 3.1.2). SA_{LAS} are calculated using inputs averaged every 60 and 10 min (Section 4.3).

The two IOP days have different radiation regimes (clear/cloudy), lateral velocity standard deviation (σ_v, Figure 5a), friction velocities (u_*, Figure 5b), intensity of instability (Figure 5c), and timing of unstable periods (Table 2). On IOP-1, the mean wind direction (216.9°, Figure 5h) is in the dominant wind sector for London (Kotthaus & Grimmond, 2014b), whereas on the clear IOP-2 day (130.4°) it comes increasingly from the east.

With more unstable conditions during IOP-2, the SA_{LAS} are smaller (60%: 0.3–3.3 km^2, 10 min) than on IOP-1 (60%: 2.7–7.4 km^2) (Figure 5f). Wind speed and direction relative to the path orientation also plays a role (Figure 7a,b). The SA_{LAS} is larger on IOP-1 when the wind direction is consistently perpendicular to the path (Figure 5h) than when it is along (IOP-2 clear day: 0600–0700; 1300–1400). Wind speeds are higher during IOP-1 (Figure 5g) with differences up to ~5 ms^{-1} in the afternoon (Figure 7a), accompanied by an increasing u_* (Figure 5b).

The largest SA_{LAS} weights (per 16-m^2 pixel) per time period indicate where the maximum influence point is (i.e., the pixel with the largest SA weight). These maxima are aligned with the wind direction and vary little through the day, with IOP-1 more similar across all hours (Figure 7).

There is dynamic feedback between the SA_{LAS} position and surface characteristics influencing the observations. Generally, with larger roughness elements the aerodynamic parameters increase (z_0, Figure 5d, and z_d) and the effective measurement height (z_f) decreases (Figure 5e). However, z_f is similar over both days (10 min: IOP-2: 65.0–76.0 m; IOP-1: 63.4–73.5 m); and the surface characteristics are similar (Figure 7c,d) despite the different radiation regimes. With the SA aligned along the path (e.g., 1400, Figure 7a,b) the smaller SA is proportionally more impacted by the City of London high-rise buildings ($z_f = 66$ m *cf.* >70 m). At the same time, the proportions of vegetation and water become smaller (Figure 7c,d). On

TABLE 2 Two days when observed (obs) incoming short-wave radiation ($K_{\downarrow}$) and turbulent sensible heat flux (Q_H) are compared to UKV outputs ($K_{\downarrow,UKV}$: 15-min mean for :01 → :15 each hour; $Q_{H,UKV}$: instantaneous on the hour), for times ending (UTC) with units W m^{-2}.

			IOP-1		IOP-2	
		Averaging period (min)	$K\downarrow$	Q_H	$K\downarrow$	Q_H
(a) Obs	*n* data points	1	796	718	740	730
		5	160	144	148	146
		10	80	72	74	73
		60	13	12	13	12
	Mean	1	229.0	122.1	535.3	223.1
	Mean SD	10	40.1	22	19.6	61.5
	Max SD	10	237.2	59.3	63.9	134.3
		when (*UTC*)	1330	1330	1410	1300
	Min SD	10	0.4	3.9	1.7	13.2
		when (*UTC*)	0530	0650	1140	1730
(b) $UKV_{surface}$	Mean		151.2	73.8	578.8	229.9
(c) Obs – $UKV_{surface}$	MAE	1	144.2	49.4	32.0	76.3
		5	125.8	45.8	32.6	73.3
		10	123.3	42.2	44.3	76.6
		60	114.3	48.1	75.3	41.5
	MBE	1	134.8	49.4	−29.0	−9.6
		5	114.7	45.8	−28.3	−13.8
		10	111.0	42.4	−41.2	−16.5
		60	95.7	47.7	−32.3	−19.8
	Max BE	1	656.6	124.5	−86.6	154.2
		when (*UTC*)	1215	1500	0915	0800
		5	466.7	127.6	−104.2	137.6
		when (*UTC*)	1215	1500	0915	0900
		10	345.1	110.9	−133.7	133.3
		when (*UTC*)	1215	1500	1415	1000
		60	282.4	115.2	−134	−95.8
		when (*UTC*)	1515	1400	0815	1400

		IOP-1		IOP-2	
		Q_H	***when* (*UTC*)**	Q_H	***when* (*UTC*)**
(d) $UKV_{surface} - UKV_{level}$	MAE	5.8		6.6	
	Max AE	15.1	1300	15.6	0700
	Min AE	1.0	1200	0.3	1300
(e) $UKV_{surface} - UKV_{centre}$	MAE	4.4		6.5	
	Max AE	12.2	1300	13.5	1100
	Min AE	0.2	1400	0.3	0700

Note: The observation periods are from 0525 to 1840 UTC on IOP-1 (a cloudy day) and from 0541 to 1800 UTC on IOP-2 (a clear day). Metrics are defined in Table SM.7. UKV fluxes are compared from three outputs: (c) *surface* is the surface flux weighted by SA_{LAS} (Section 4.3), (d) *level* is the UKV grid box vertically closest to z_f and closest to the centre of the LAS path, and (e) *centre* is the grid box closest to the LAS path centre but at z_{ES}.

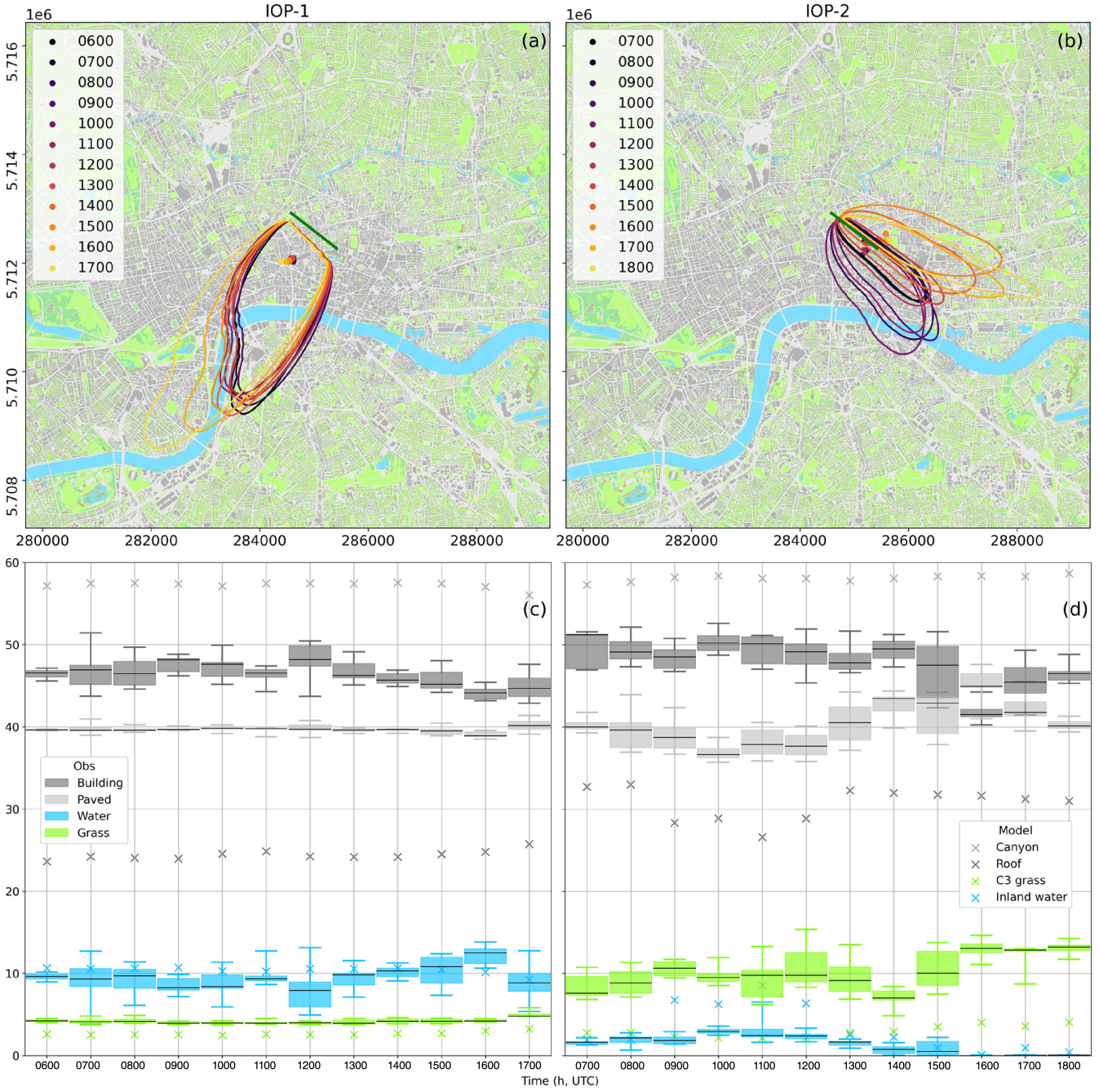

FIGURE 7 Hourly characteristics of the 60% LAS source areas on (a,c) IOP-1 and (b,d) IOP-2 showing (a,b) location (colour polygons) and IMU-BCT (Figure 4) LAS path (green) and the maximum source area weight pixel (dot) on a land cover map (as Figure 4a); and (c,d) box plots of land cover proportions (%) within each hour based on the 10-min SA weights [median (line), interquartile range, minimum/maximum (whiskers)] and SA weighted UKV grid boxes land cover (crosses, Section 4.3). [Colour figure can be viewed at wileyonlinelibrary.com]

IOP-1, meteorological consistency results in little diurnal variation in SA_{LAS} surface characteristics.

The SA_{LAS} built land cover proportions are consistently > 80% (Figure 7a), as found previously in this area between December 2014 and December 2015 (Crawford et al., 2017) with the rest grass and water (Figure 7). Overall, on IOP-1, water accounts for ~10% of the SA_{LAS} (*cf.* IOP-2, <5%). On IOP-1, all SA_{LAS} include a portion over the River Thames (Figure 7a), and the land cover includes more water than on IOP-2 (Figure 7c,d: 8%–13%, *cf.* 0%–3%) where grass surfaces are more important (7%–13%, *cf.* 4%–5%).

Variability of the plan area of buildings, derived from the six 10-min SA_{LAS} per hour, is small with the IOP-2 IQR at 6.3% at 1500 (Figure 7d). SA_{LAS} land cover variation is generally larger on IOP-2 (*cf.* IOP-1) where the average IQR across all hours for grass is 2.1% (0.3%) for IOP-2 (IOP-1), 2.3% (0.4%) for paved, and 2.6% (1.8%) for building. This is linked to the larger range in wind

direction (Figure 5h), resulting in some SA_{LAS} extending beyond the City of London local authority into areas with more grass (Figure 7).

Different UKV grid boxes are sampled depending on the situation (Table SM.6). The largest SA_{LAS} overlap with a UKV model grid box (Figure 7) on IOP-1 is for grid box E (Figure 4a). On IOP-2, grid box A (with the largest proportion of the LAS path) and grid box B are both covered with similar weights.

Assessment of the horizontal consistency (Section 2.1) involves dynamically weighting the UKV grid boxes using the SA_{LAS} (Section 4.3), using the model land cover fractions (Figure 4b) and the high-resolution DSM data (Figure 7c,d). The UKV built land cover proportions differ on both days from the SA_{LAS} results, with UKV 'canyon' (paved) proportions larger than 'roof' (building), which is the inverse to that found in the observations. The SA_{LAS} building maxima are slightly larger for IOP-2 (53%) than IOP-1 (51%), whereas a greater contrast occurs in the paved maxima (48% and 41%, respectively). However, the UKV (crosses, Figure 7c,d) building land cover (IOP-2: 33%, IOP-1: 26%) is less than paved (IOP-2: 59%, IOP-1: 58%). Overall, the total built (building + paved) fractions are in better agreement, with both IOP-1 (UKV 82%–83%; LAS 82%–92%) and IOP-2 (UKV: 85%–91%; LAS: 82%–95%) values mostly overlapping.

5.2 | Validity of short averaging-period fluxes

Statistically stable results of $Q_{H,LAS}$ can be obtained for very short averaging periods, as low as 1 min (Beyrich et al., 2021). The validity of this study's shortest averaging times of $Q_{H,LAS}$ is investigated by comparing 1-min means to the 2-, 3-, 5-, 10-,15-, 20-, 30- and 60-min fluxes.

The absolute difference between median (Figure 8, white) and 5-min fluxes (values given in Table SM.3) for IOP-1 are smaller for 1-, 2-, 3-min fluxes (means <4%). Differences increase with longer flux averaging periods of 10, 15, 20, 30 and 60 min (means: 5%–27%). The IOP-2 differences are relatively smaller for 1–20-min fluxes (means <6%) compared to the 30- and 60-min fluxes (means 10%–20%). The differences start to become more evident at averaging periods of 10 min and longer (Figure 8).

There are instances of $Q_{H,LAS}$ where the response to changes of incoming short-wave radiation ($K_{\downarrow}$) is evident. Examination of the length of delay between a $K_{\downarrow}$ change and a $Q_{H,LAS}$ response is investigated by finding the peak Pearson correlation coefficient between the independently measured $K_{\downarrow}$ and $Q_{H,LAS}$, for different time lags Δt, for $-100\,\text{min} \geq \Delta t \geq 100\,\text{min}$ using 1-min increments (Figure SM.2). Here, the maximum correlation is found when $Q_{H,LAS}$ is shifted by −9 min on IOP-1 and −13 min on IOP-2. This result can therefore explain the differences seen between the 1-, 2-, 3-, 5-min flux averages and the 10-min or more average fluxes. Changes in wind direction and radiation during IOP-1 and 2 also influence these results; the flatter curve (Figure SM.2) during IOP-2 indicates that the response in Q_H is less pronounced, *cf.* IOP-1, due to the clear sky conditions and the radiation site (KSSW, Figure 4) not being within the observation SAs.

To explore this further, we consider one hour with relatively large intra 10-min $Q_{H,LAS}$ variability compared to other hours on this day (IOP-2, 1330 to 1430 UTC, see Figure SM.3b). We investigate the $Q_{H,LAS}$ outlier of $374\,\text{W m}^{-2}$ at 1410 (Figure SM.3a), that follows a peak in the 10-min $K_{\downarrow}$ mean of $782\,\text{W m}^{-2}$ at 1400, with a subsequent drop to $600\,\text{W m}^{-2}$ at 1410. The 10-min $Q_{H,LAS}$ decreases to $241\,\text{W m}^{-2}$ at 1420. The 1-min $Q_{H,LAS}$ shows a similar response, with a peak of $460\,\text{W m}^{-2}$ at 1405, which falls to $245\,\text{W m}^{-2}$ at 1411.

Thus, the short averaging period observations have a physical response to radiation changes in $Q_{H,LAS}$ which the NWP model (at 1.5 km resolution) is not capable of capturing. In the UKV evaluation context, 10-min averages are likely more useful than those no longer than 5 min, despite these capturing physical processes. However, these shorter periods are likely to be more useful as NWP modelling moves to higher resolutions, such as the 100-m resolution configuration of the UK Met Office's Unified Model (UM100; Lean et al., 2019).

Given the evidence for horizontal convective rolls in the UM100 (Hall et al., 2024), and that over short periods a sensor could observe net up-/downdrafts in the highest weighted portion of the path (i.e., centre), large fluctuations in 1-min $Q_{H,LAS}$ are not unexpected (Figure 9). With multiple sensors in close vicinity, we would expect to observe both updrafts and downdrafts at the same time, but in different locations. The variability is greater on the clear IOP-2 (Figure 9b) day than on the cloudy IOP-1 (Figure 9a), as the latter will have less strong thermally induced updrafts.

5.3 | Variability of observed sensible heat and incoming short-wave radiation fluxes

Before assessing the temporal consistency (Section 2.3) between $Q_{H,LAS}$ and $Q_{H,UKV}$, we first look at $Q_{H,LAS}$ and observed $K_{\downarrow}$, as the temporal pattern of Q_H is largely dependent on $K_{\downarrow}$. The 1-min $Q_{H,LAS}$ follows $K_{\downarrow}$ with an expected overall phase difference, but many of the impacts of the minute-to-minute fluctuations in $K_{\downarrow}$ can be seen in $Q_{H,LAS}$ (Figure 9). Similarly, as expected, both fluxes are larger on the clear day (IOP-2) than during IOP-1 (Table 2). Intra 10-min $Q_{H,LAS}$ variability is greatest when

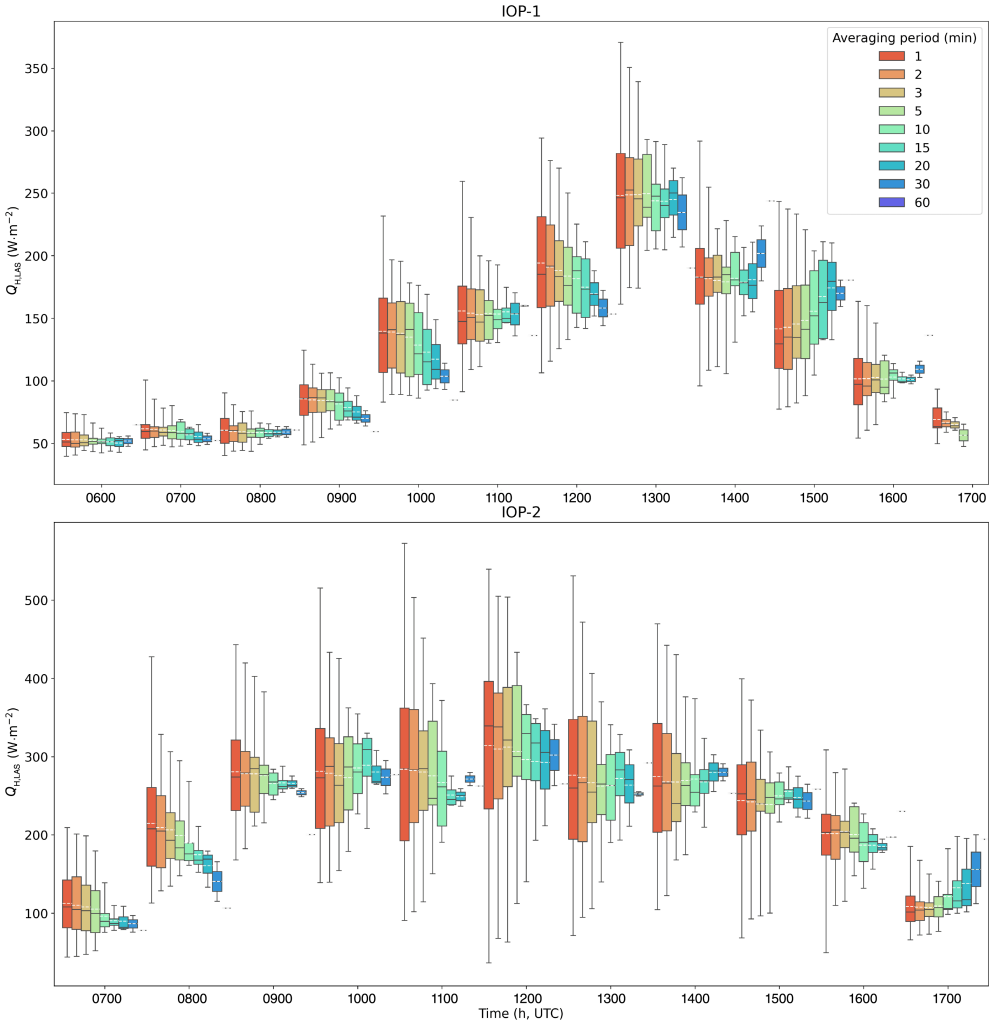

FIGURE 8 Hourly $Q_{H,LAS}$ variability of fluxes determined for 1-, 2-, 3-, 5-, 10-, 15-, 20-, 30-, 60-min averaging periods on the two IOP days, showing IQR (box), median (black), mean (white dashed), and min/max values (whiskers). [Colour figure can be viewed at wileyonlinelibrary.com]

the fluxes are larger, so occur during the middle of day (*cf.* sunrise/sunset) and on IOP-2 (*cf.* IOP-1) (Figure 9). At night (i.e., $K_{\downarrow} = 0\,\mathrm{W\,m^{-2}}$) in this densely built-up study area, Q_H will rarely drop below $0\,\mathrm{W\,m^{-2}}$ due to the release of long-wave radiation, anthropogenic and storage heat fluxes. Therefore, during morning and evening transition periods when $K_{\downarrow}$ is small (e.g., cloudy IOP-1) it is physically reasonable for Q_H to exceed $K_{\downarrow}$.

To assess the impact of SA surface cover composition we analyse midday (1000–1400) Q_H normalised by $K_{\downarrow}$ as a function of the built (paved + building) land cover for both observations and model. As the variation in the built proportion observed is small (Figure 7, between 45% and 50%), so is the impact on $Q_H/K_{\downarrow}$ (Figure SM.4). Although Crawford et al. (2017) found $Q_{H,LAS}/Q^*$ varies (see their Figure 10), their SA_{LAS} building fractions vary between ~25% and 50% in their analysis of a longer period and multiple LAS paths.

The SA_{LAS} surface characteristics impact the $Q_{H,LAS}$ calculated via the aerodynamic roughness parameters and therefore z_f. Two cases are compared with SA_{LAS} calculated for each hour ($Q^{hour}_{H,LAS}$) and each 10-min ($Q^{10\ min}_{H,LAS}$) period (Figure 10, median values). There is a 15.2% difference ($[x_{10min} - x_{hour}]/x_{hour}$) in z_f, with a ~0.82 slope with $Q_{H,LAS}$, but as eq. F.22 (Table 1) indicates, this is not a linear relation. This suggests changes in surface characteristics are sufficient to require SA_{LAS} to be recalculated when deriving $Q_{H,LAS}$. The intra-period variability (i.e., IQR lines, Figure 10) is larger on IOP-2. When the SA_{LAS} are more dynamic (Figure 7), such as caused by changing wind directions (Figure 5h), it would be beneficial for the derivation of $Q_{H,LAS}$ to calculate SA_{LAS} more frequently. Time of day does not seem to be as critical to IQR size in these two days.

5.4 | Impact of new source area methodology

The new method to determine SA_{LAS} uses varying aerodynamic roughness parameters (z_0 and z_d) along the LAS path derived from multiple unique SA_{EC} (Figure 2a,d), which builds upon Crawford et al.'s (2017) methods with one SA_{EC} applied along the path. To assess this, we quantitatively compare Kanda et al. (2013) hourly roughness

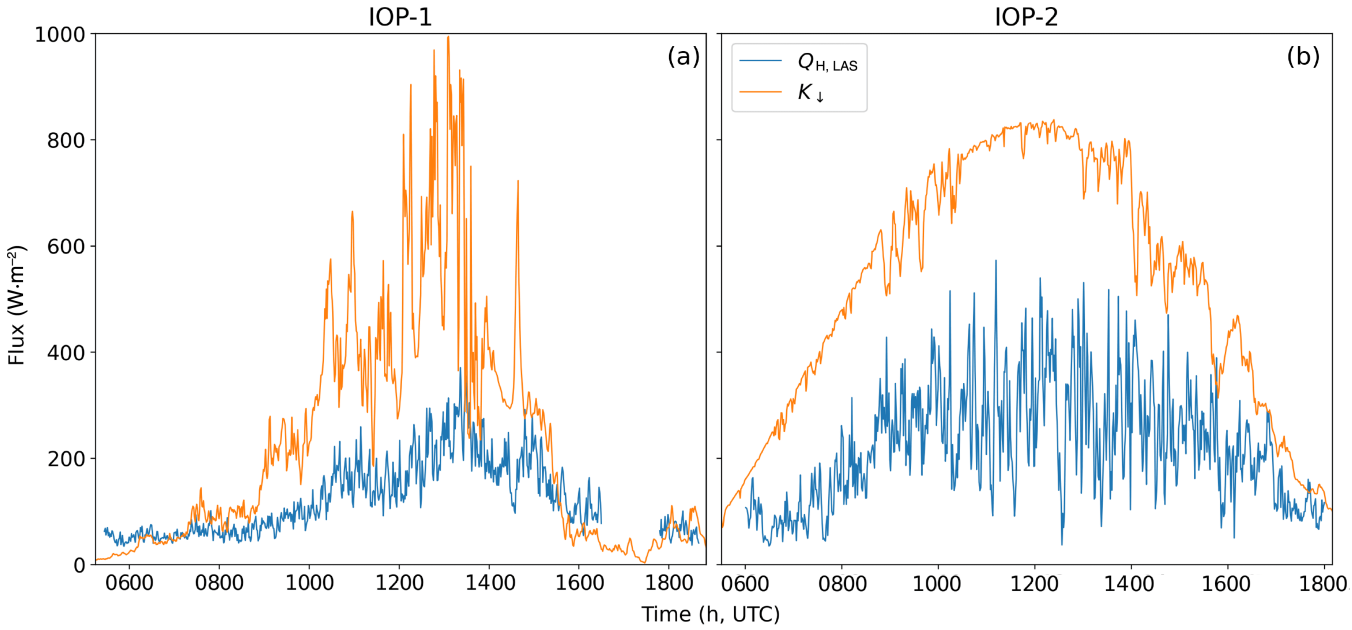

FIGURE 9 Observations (1-min mean) of $Q_{H,LAS}$ and incoming short-wave radiation ($K_\downarrow$) on (a) a cloudy (IOP-1) and (b) a clear day (IOP-2). [Colour figure can be viewed at wileyonlinelibrary.com]

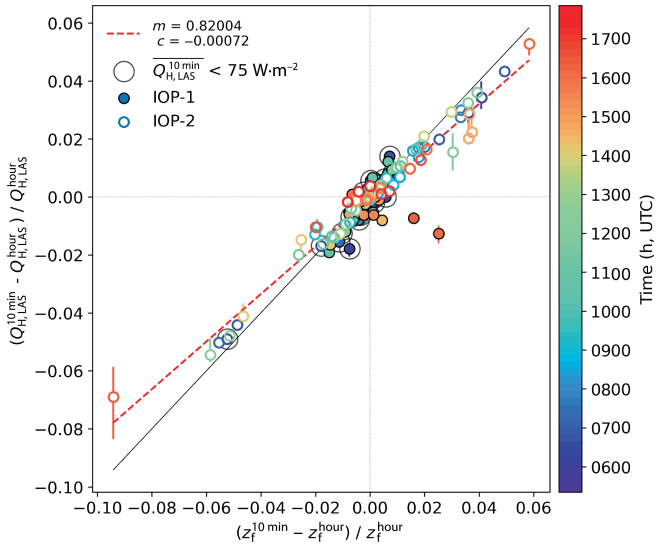

FIGURE 10 The impact of frequency of SA_{LAS} calculation (10 min vs. one hour) on effective measurement height (z_f) and $Q_{H,LAS}$ by time of day, shown as normalised differences $(x^{10\,min} - x^{hour})/x^{hour}$ with $Q_{H,LAS} < 75\,W\,m^{-2}$ (large circle) indicated. Variability of 10-min intervals derived from 1-min data with median (markers) and IQR (lines) shown. The gradient (m) and y-axis intercept (c) for the best-fit line (linear regression) are listed. [Colour figure can be viewed at wileyonlinelibrary.com]

parameters for the whole SA_{LAS} (Figure 5d,e and SM.5) every 50 m along the path (Figure 2d): (1) constant parameters (SA_{LAS}^{old}) and (2) varying parameters (Section 3.1.1, SA_{LAS}).

The percentage difference in z_d between SA_{LAS} and SA_{LAS}^{old} ($[SA_{LAS} - SA_{LAS}^{old}]/SA_{LAS}$) is above 10% for eight out of the 24 hours assessed over the two IOP days (z_0: 16 of 24 h).

Changing z_d causes a 0%–5% difference in the LAS effective measurement height (z_f, eq. F.9), which is critical to obtaining $Q_{H,LAS}$ (Figure 10) which also changes by 0%–5%. Hence, the extra computational expense of the new SA_{LAS} methodology is warranted.

6 | MODEL–OBSERVATION COMPARISON RESULTS

6.1 | Comparison of incoming short-wave radiation $K_\downarrow$

Although our focus is on the turbulent sensible heat flux (Q_H), if the short-wave radiative forcing is incorrectly modelled, Q_H should be impacted, otherwise a 'well' modelled Q_H would be correct for the wrong reasons. Hence, we first establish the level of agreement between observed and modelled short-wave radiative forcing on each day.

6.1.1 | Horizontal consistency: $K_\downarrow$

The radiation measurements are compared to UKV grid box D (Figure 4) where the radiation sensor at site KSSW is located. Given the wind direction on cloudy IOP-1

(Figure 5), the KSSW radiation sensor is located in all SA_{LAS} (Figure 7a), but not on clear IOP-2 (Figure 7b) when the spatial variations of $K_{\downarrow}$ are small. For the latter, we can assume the impact from using the radiation observation site from outside the SA_{LAS} to be minimal, and not impacting interpretation of the larger spatio-temporal variability seen on cloudy IOP-1.

There is $K_{\downarrow,UKV}$ variability between the nine grid boxes which the SA_{LAS} at some point in time covers. On IOP-1, based on a 60-min time step, the minimal SA_{LAS} weighting (Section 3.1.1) for an individual UKV grid box is 1% and the maximum is 79%. For IOP-2, this range of weights is 1%–87% (Table SM.6).

6.1.2 | Temporal consistency: $K_{\downarrow}$

On IOP-1, $K_{\downarrow,UKV}$ is underpredicted (Figure 11a), with an MBE (Section 4.3) of 96 W m^{-2} (using 60-min observations, Table 2c). Hourly $K_{\downarrow,UKV}$ output (Figure 11a,b) are 15-min means for the first part of an hour (i.e., :01 → :15, plotted at ending time). The observed and modelled $K_{\downarrow}$ peaks both occur at around 1300, but the smaller observed peak at 1100 is not as well represented in the UKV. On IOP-2, the model has no forecasted cloud, and the model's magnitude and timing of $K_{\downarrow}$ agree with observations (Figure 11b).

The 'hourly' UKV radiative fluxes are a mean of the ~1-min model timesteps over 15 min, whereas the observations are the mean of 20-s samples, with comparisons made for different lengths of observation periods. Under cloudy conditions (IOP-1), it is difficult for any model to get the incoming radiation fluxes correct, given the patchy nature of clouds not necessarily simulated to occur within the correct individual grid box, whereas, on a clear-sky day, with cloud-caused variations irrelevant, it is much easier for a model to get the radiation physics correct. Hence, on IOP-1, the best performance (MAE 114 W m^{-2}) occurs for the longest averaging period (60 min). Whereas on the clear day (IOP-2) temporal changes of solar geometry become critical, so the best performance (MAE 32 W m^{-2}, Table 2c) is for the shortest averaging period (1 min). Conversely, the poorest performance for IOP-1 occurs when the shortest time-averaging is analysed (1 min Table 2).

6.2 | Comparison of sensible heat fluxes

Prior urban evaluations of MetUM-JULES at the km scale (i.e., online) have been undertaken in London (Best, 2005; Bohnenstengel et al., 2011; Warren et al., 2018), Singapore (Simón-Moral et al., 2020), Seoul (Wie et al., 2020) and Berlin (Fenner et al., 2024). These evaluations have used observations of air temperature (Best, 2005; Bohnenstengel et al., 2011; Simón-Moral et al., 2020; Wie et al., 2020), $Q_{H,EC}$ (Simón-Moral et al., 2020; Warren et al., 2018, Fenner et al. 2024), $Q_{E,EC}$ (Simón-Moral et al., 2020, Fenner et al. 2024), and ceilometer lidar profiles of attenuated backscatter (Warren et al., 2018). Additionally, offline (i.e., uncoupled from MetUM) evaluations using observed energy balance fluxes (e.g., $Q_{H,EC}$, $Q_{E,EC}$) have been undertaken for the central business district area in London (Hertwig et al., 2020) and Mexico City (Best et al., 2006; Porson et al., 2010a, 2010b), industrial Vancouver (Best & Grimmond, 2013; Grimmond, Blackett, Best, Barlow, et al., 2010), and residential Melbourne (Grimmond, Blackett, Best, Baik, et al., 2010; Lipson et al., 2024). These provide a context of expectations, for our results.

6.2.1 | Horizontal consistency: Q_H

Our investigation of horizontal consistency between LAS and the UKV (Section 2.1) finds the UKV canyon plan area to be assigned too large a value based on assessment with the reference data (Section 4.1). Hertwig et al. (2020) demonstrate that the canyon to roof proportion critically effects the timing of MORUSES modelled Q_H. Larger canyon proportions cause a longer delay in the rise of Q_H, due to the canyon's larger thermal inertia in comparison to the roof, which in turn responds quickly to changes in the radiative forcing.

$Q_{H,UKV}$ varies between grid boxes on both days, unlike $K_{\downarrow}$ on IOP-2 (Figure 11). The MAE between the surface $Q_{H,UKV}$ for the centre of the path grid box (A, Figure 4a) and the grid box weighted is 4 W m^{-2} (Table 2e). This is smaller than the differences seen between $Q_{H,UKV}$ at z_{ES} and on model levels (Table 2).

6.2.2 | Vertical consistency: Q_H

Although Q_H at the model's effective surface is considered to be within the ISL (Figure 1a,b), the fluxes show distinct height variations (Figure 12). For model values from the grid box at the centre of the path (i.e., not weighted by SA_{LAS}), the difference between $Q_{H,UKV}$ at the effective surface, and $Q_{H,UKV}$ at the model level closest to the z_f (level at 62 m) is between 0.5 and 20 W m^{-2} on IOP-1 and 10–31 W m^{-2} on IOP-2 (Table 2d). On IOP-2, the largest percentage difference over the analysis period (Table 2d) between the SA_{LAS}-weighted $Q_{H,UKV}$ at z_{ES} ($UKV_{surface}$, red) and UKV_{level} (blue, Figure 11c,d) is 31%, (IOP-1: 24%).

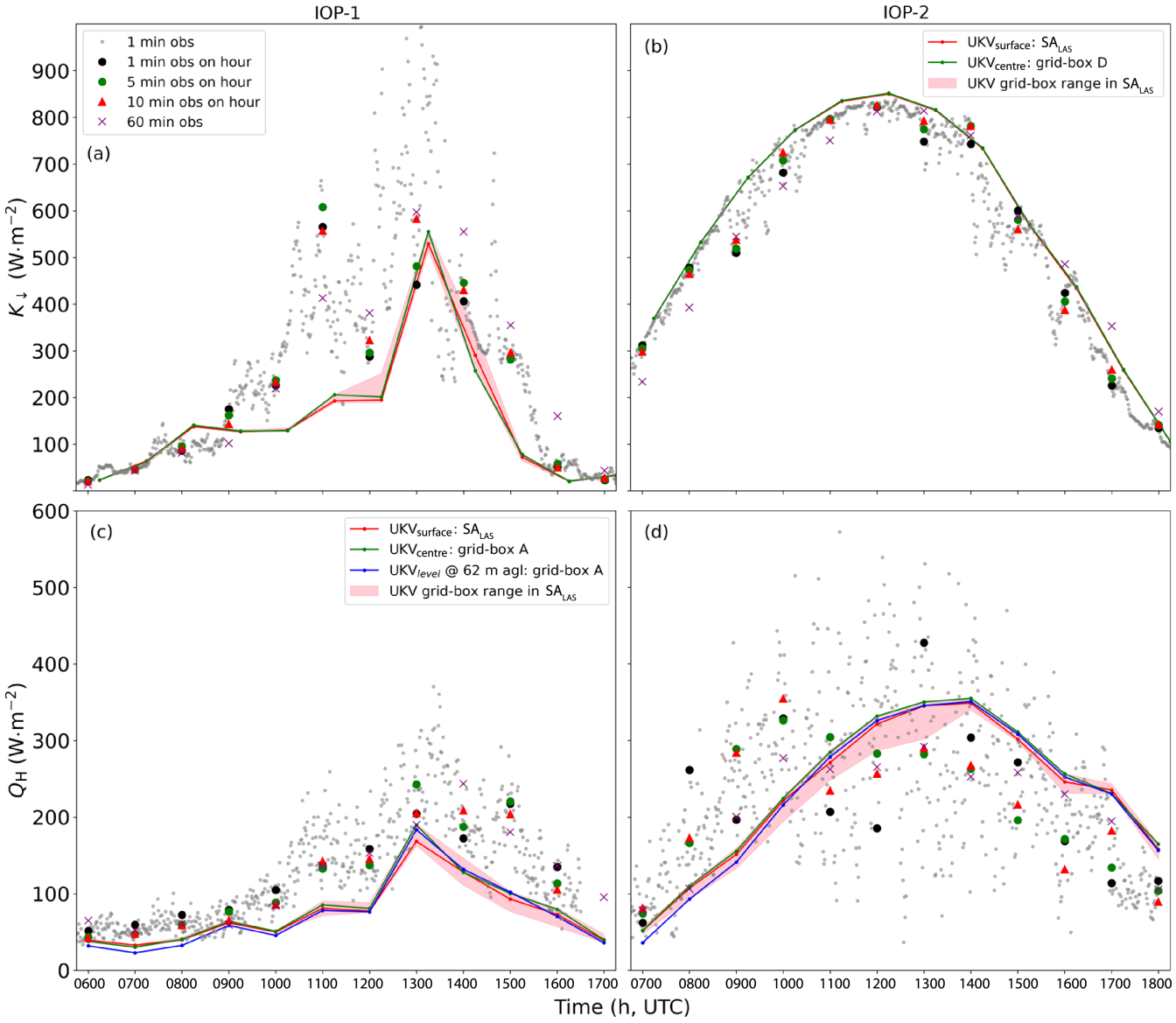

FIGURE 11 Observations for different averaging periods (symbols) and UKV model output (lines) for (a,c) IOP-1 and (b,d) IOP-2, showing (a,b) $K_{\downarrow}$ (UKV: 15-min mean each hour from :00 → :15, plotted at time ending) and (c,d) Q_H (UKV: sample time) surface fluxes SA_{LAS}-weighted (weighted mean: red line; spatial variability from the UKV grid box range: red shading) and for one grid box for (a,b) surface $K_{\downarrow,UKV}$ (green, 'D' Figure 4) with radiation site KSSW and (c,d) for $Q_{H,UKV}$ ('A', Figure 4) for LAS path centre at the surface (green) and at 62 m above ground level (a.g.l.; blue). [Colour figure can be viewed at wileyonlinelibrary.com]

Comparing the $Q_{H,UKV}$ between the surface flux and the boundary-layer flux (model level at 62 m), the MAE for IOP-1 is smaller than for IOP-2. As the largest difference is between the surface and the 62 m level (16 W m^{-2}, Table 2d), it is beneficial to analyse both surface and model level Q_H.

6.2.3 | Temporal consistency: Q_H

During IOP-2, a phase shift is visible between $Q_{H,LAS}$ and $Q_{H,UKV}$ (Figure 11), where the increase in Q_H occurs sooner in the LAS observations than the model, which has higher afternoon fluxes. This phase delay in MORUSES Q_H, over urban areas, has been previously reported in the literature (Hertwig et al., 2020; King, 2015; Porson et al., 2010a, 2010b; Warren et al., 2018). This has been attributed to building and other urban material's high inertia, dampening the diurnal amplitude of surface temperature (Harman & Belcher, 2006).

On IOP-1, timing of the $Q_{H,UKV}$ peak is consistent with the $K_{\downarrow,UKV}$ peak. The $Q_{H,UKV}$ is underpredicted relative to $Q_{H,LAS}$ in all four cases evaluated (Figure 11): (1) UKV$_{surface}$ (red), (2) UKV$_{centre}$ (green), (3) UKV$_{level}$ (blue) and (4) the

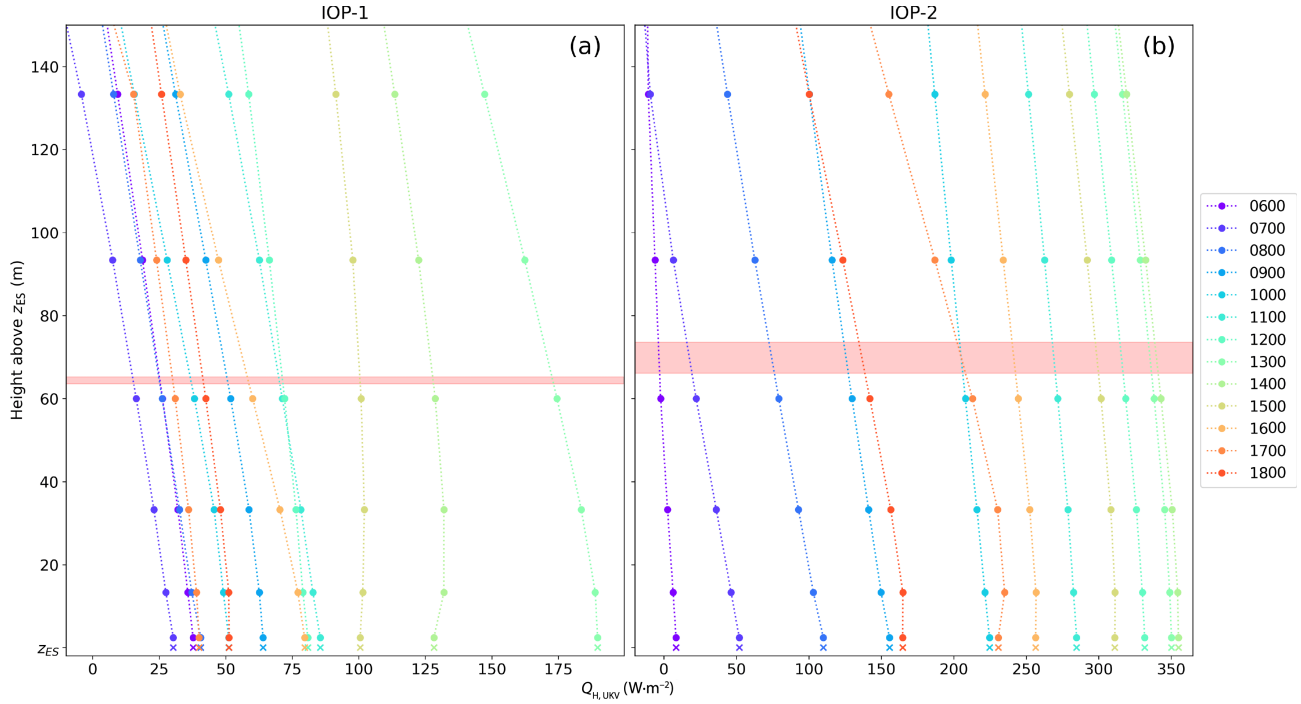

FIGURE 12 Central grid box of LAS path (A, Figure 4) $Q_{H,UKV}$ from surface (*crosses*) and other model levels (dots) for different times (h UTC, colour) on (a) IOP-1 and (b) IOP-2 (Section 4) with calculated LAS-z_f (red shading, Section 3.1.1). [Colour figure can be viewed at wileyonlinelibrary.com]

range of grid boxes the SA_{LAS} covers (pink shading). This is consistent with the underprediction of $K_{\downarrow}$ for this day (Section 6.1).

The largest BE occurs between the 1-min mean $Q_{H,LAS}$ and $UKV_{surface}$, (154 W m^{-2}, Table 2c). On IOP-2, the UKV grid box range is within the LAS observational temporal variability (all intervals). The largest difference is in the morning for 1-min averages (Table 2c).

Figure 13 shows how model performance varies on both days using four $Q_{H,LAS}$ averaging periods. For IOP-2 (solid lines), there is a change in the model bias from morning to afternoon, showing the phase delay in MORUSES (Figure 11d). This is consistently seen across all four LAS averaging periods. At 1700 the normalised model bias exceeds 1, indicating an overprediction of the flux of more than 100%. On IOP-1 (dotted lines), the model always underpredicts the flux, with no morning–afternoon change in model bias. IOP-1's reduced radiative forcing (*cf.* IOP-2) causes a subdued thermal response. Model performance (Figure 13) is best using 10-min averages on IOP-1, and 60-min averages on IOP-2. For both days, the poorest performance occurs when using 1-min averages.

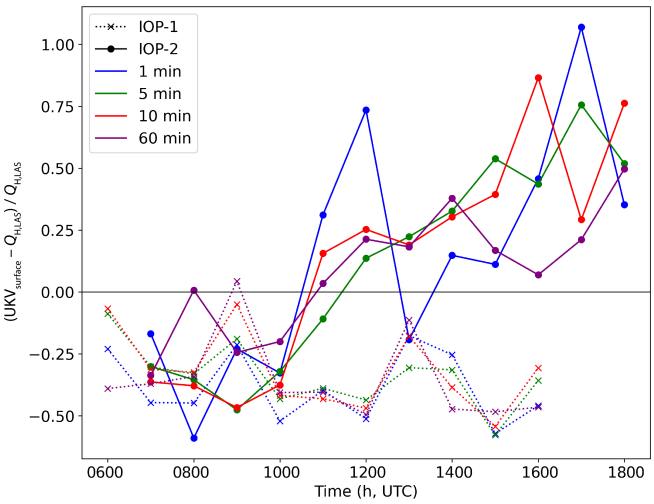

FIGURE 13 Difference between observed ($Q_{H,LAS}$) and modelled ($Q_{H,UKV}$; instantaneous output on the hour at the 'surface' for grid boxes included in the source area analysis $UKV_{surface}$) for four averaging periods (colour), normalised by $Q_{H,LAS}$ for IOP-2 (solid) and IOP-1 (dotted). [Colour figure can be viewed at wileyonlinelibrary.com]

7 | DISCUSSION AND CONCLUSIONS

Here a methodology to evaluate NWP of turbulent sensible heat flux (Q_H) using LAS observations is developed, which considers the impacts of surface variability. Our current aim is not to assess the performance of the NWP model, but to identify what needs to be considered and demonstrate how to achieve these steps.

Our approach is intended to have more spatially consistent scales with NWP models, than more common EC

fluxes with smaller SAs (Ward et al., 2014). Additionally, LAS fluxes can be calculated for shorter time periods (1, 5, 10 min, and longer) than for EC fluxes (30, 60 min), allowing comparison with a model's instantaneous sample values. Consideration is given to the changing effective measurement height as the SA_{LAS} varies in response to meteorology and surface characteristics. LAS sensor deployment has advantages in urban areas of not requiring towers to be installed, which can be logistically challenging for EC in cities.

Assumptions and limitations involved in obtaining $Q_{H,LAS}$, especially when measured over complex terrain, do exist and have been explored in previous literature (Ward, 2017). One of the current limitations in the literature addressing SA_{LAS} calculation methods are improved upon in this work. Notably, roughness parameters now vary along the path (previously assumed to be constant, e.g., Crawford et al., 2017).

The methodology considers both spatial (horizontal, vertical) and temporal aspects of modelled and observed fluxes:

Horizontal

- The SA_{LAS} uses the LAS path-weighting function to combine individual point footprints spaced at intervals (here 50 m, Section 3.1.1) along the path.
- As changes in surface characteristics within the SA_{LAS} impact the magnitude of $Q_{H,LAS}$ (Figure 10), we suggest it is appropriate to calculate SAs more frequently using the meteorological inputs observed over short time intervals (here: 10 min). If the dynamically changing SA_{LAS} positions are more variable than the cases demonstrated here, more frequent SA_{LAS} calculations may be warranted (<10 min), as the derived surface parameters cause the largest variability in the iteratively derived $Q_{H,LAS}$ (Figure 10). Here, time of day is not found to be a critical for selecting the SA_{LAS} calculation frequency.
- For model-observation horizontal spatial consistency, the appropriate NWP grid box fluxes are weighted using the SA_{LAS} accumulated weight per grid box (e.g., Figure 2).
- NWP land cover data can be both coarse and dated, making model surface parameters less representative of the real world and biasing model skill unnecessarily. Here the Met Office's NWP model (UKV) has smaller areas of vegetation, and more paved areas compared to buildings, than the SA_{LAS} dataset (Figure 7). However, total built proportions (building + paved) are in better agreement.

Vertical

- For vertical consistency, the closest model level to observation z_f is identified and used for each period compared.
- Additionally, the NWP surface flux is compared to the $Q_{H,LAS}$. Both are considered to be in the inertial sublayer (ISL).

Temporal

- $Q_{H,LAS}$ can be obtained for very short periods (e.g., 1 min) unlike EC over surfaces with large roughness elements (Haugen, 1973), with SA_{LAS} changing over consecutive time periods because of meteorological conditions and surface characteristics within the SA_{LAS}.
- 1-min mean $Q_{H,LAS}$ are shown to respond in a manner consistent to observed 1-min mean $K_{\downarrow}$ (Figure 9). This suggests that with higher-resolution modelling, that is turbulence permitting, the proposed methodology will have increasing utility.

The method is demonstrated for two days with contrasting cloud cover using one LAS path in central London and the UKV NWP model. The following conclusions are drawn:

- During evaluation, when $Q_{H,LAS}$ fluxes (1 min, 5 min) with a similar frequency to the UKV's time step (order 1 min) are used, the model performance is poorer than when longer $Q_{H,LAS}$ means (e.g., 10 min, 60 min) are used. Poorest performance occurs using 1-min $Q_{H,LAS}$ despite this being the closest time interval to the model. This is attributed to the UKV not explicitly representing turbulence. This conclusion is likely to hold for other models with similar constraints. Although land cover varies between grid boxes, the parametrised UKV surface fluxes are similar (*cf.* observations) as the turbulence is not explicitly resolved. Hence, if the UKV correctly predicts an observed extreme within a given hour, it is not representative of a 60-min period.
- It is important to consider at what vertical level to evaluate, possibly more important than which grid boxes are included, when the surface cover variations are small. With little variability in built proportion, the $Q_H/K_{\downarrow}$ values are similar, suggesting reasonably consistent fetch. The small land cover variability results in small differences in the model evaluation results between four different $Q_{H,UKV}$ flux values (Section 6.2). Larger differences occur between $Q_{H,UKV}$ surface and the model level closest to LAS-z_f, although both are within the model's ISL or constant flux layer.

Application of this methodology will allow longer-term model evaluation using LAS to be conducted, and conclusions about the model's performance to be made. In addition, the LAS observations can be used to evaluate higher-resolution models (e.g., Boutle et al., 2016; Lean et al., 2019), where surface variability should be better captured through providing more detailed land cover datasets.

ACKNOWLEDGEMENTS

We thank the two reviewers of this paper for their thorough, constructive and helpful comments which have helped us improve the paper. The authors acknowledge the funding and support from the NERC Case studentship (NE/N008103/1) and ERC urbisphere (855005). We thank Ben Crawford (University of Denver) for his support and guidance on his observation SA and data processing methods; Islington Borough Council, City of London, and King's College London for providing site access; and other members of the Micromet research group (University of Reading) for their valuable contributions to fieldwork, discussion, and meteorological data processing. We thank Met Office staff, Chloe Eagle, Khalid Mahmood and James Manners, for their help with obtaining and interpreting the model data.

FUNDING INFORMATION

NERC Case studentship (NE/N008103/1) and ERC urbisphere (855005).

DATA AVAILABILITY STATEMENT

See Table SM.1.

ORCID

Beth Saunders https://orcid.org/0000-0002-2788-6134
Sue Grimmond https://orcid.org/0000-0002-3166-9415
Denise Hertwig https://orcid.org/0000-0002-2483-2675
Sylvia I. Bohnenstengel https://orcid.org/0000-0001-6170-5774
Humphrey W. Lean https://orcid.org/0000-0002-1274-4619
William Morrison https://orcid.org/0000-0003-1568-8226

REFERENCES

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

Ambaum, M.H.P. (2020) *Thermal physics of the atmosphere*, 2nd edition. Amsterdam, Netherlands: Elsevier, p. 268.

Andreas, E.L. (1988) Estimating Cn^2 over snow and sea ice from meteorological data. *Journal of the Optical Society of America A*, 5, 481. Available from: https://doi.org/10.1364/josaa.5.000481

Barlow, J., Best, M., Bohnenstengel, S.I., Clark, P., Grimmond, S., Lean, H. et al. (2017) Developing a research strategy to better understand, observe, and simulate urban atmospheric processes at kilometer to subkilometer scales. *Bulletin of the American Meteorological Society*, 98, ES261–ES264. Available from: https://doi.org/10.1175/BAMS-D-17-0106.1

Barlow, J.F., Harrison, J., Robins, A.G. & Wood, C.R. (2011) A wind-tunnel study of flow distortion at a meteorological sensor on top of the BT tower, London, UK. *Journal of Wind Engineering and Industrial Aerodynamics*, 99, 899–907. Available from: https://doi.org/10.1016/j.jweia.2011.05.001

Best, M.J. (2005) Representing urban areas within operational numerical weather prediction models. *Boundary-Layer Meteorology*, 114, 91–109. Available from: https://doi.org/10.1007/s10546-004-4834-5

Best, M.J. & Grimmond, C.S.B. (2013) Analysis of the seasonal cycle within the first international urban land-surface model comparison. *Boundary-Layer Meteorology*, 146, 421–446. https://doi.org/10.1007/s10546-012-9769-7

Best, M.J., Grimmond, C.S.B. & Villani, M.G. (2006) Evaluation of the urban tile in MOSES using surface energy balance observations. *Boundary-Layer Meteorology*, 118, 503–525. Available from: https://doi.org/10.1007/s10546-005-9025-5

Best, M.J., Pryor, M., Clark, D.B., Rooney, G.G., Essery, R., 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, 677–699. Available from: https://doi.org/10.5194/gmd-4-677-2011

Beyrich, F., De Bruin, H.A.R., Meijninger, W.M.L., Schipper, J.W. & Lohse, H. (2002) Results from one-year continuous operation of a large aperture scintillometer over a heterogeneous land surface. *Boundary-Layer Meteorology*, 105, 85–97. Available from: https://doi.org/10.1023/A:1019640014027/METRICS

Beyrich, F., & Coauthors, (2012) Towards a validation of scintillometer measurements: the LITFASS-2009 Experiment. *Boundary-Layer Meteorol*, 144, 83–112, Available from https://doi.org/10.1007/S10546-012-9715-8/METRICS

Beyrich, F., Hartogensis, O.K., de Bruin, H.A.R. & Ward, H.C. (2021) Scintillometers. In: Schröder, F. & Höhlein, K. (Eds.) *Springer handbook of atmospheric measurements*. Berlin, Germany: Springer, pp. 969–997.

Bing Maps. (2021) Bing Maps - Directions, trip planning, traffic cameras & more. https://www.bing.com/maps/?cp=51.521607~-0.096694&lvl=16.8&style=h (Accessed August 26, 2021).

Bohnenstengel, 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, 1625–1640. Available from: https://doi.org/10.1002/qj.855

Boutle, I., Lock, A. & Edwards, J. (2021) The Parametrization of Boundary Layer Processes, Unified Model Documentation Paper 024.

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, 360–371. Available from: https://doi.org/10.1002/qj.2656

Britter, R.E. & Hanna, S.R. (2003) Flow and dispersion in urban areas. *Annual Review of Fluid Mechanics*, 35, 469–496. Available from: https://doi.org/10.1146/annurev.fluid.35.101101.161147

Capel-Timms, I., Smith, S.T., Sun, T. & Grimmond, S. (2020) Dynamic anthropogenic activitieS impacting heat emissions (DASH v1.0): development and evaluation. *Geoscientific Model Development*, 13, 4891–4924. Available from: https://doi.org/10.5194/GMD-13-4891-2020

CEH. (1990) *Land cover map of Great Britain 1990*. United Kingdom: Centre for Ecology & Hydrology. https://www.ceh.ac.uk/services/landcover-map-1990

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, 701–722. Available from: https://doi.org/10.5194/gmd-4-701-2011

Crawford, B., Grimmond, C.S.B., Ward, H.C., Morrison, W. & Kotthaus, S. (2017) Spatial and temporal patterns of surface–atmosphere energy exchange in a dense urban environment using scintillometry. *Quarterly Journal of the Royal Meteorological Society*, 143, 817–833. Available from: https://doi.org/10.1002/qj.2967

De Bruin, H.A.R., Kohsiek, W. & Van Den Hurk, B.J.J.M. (1993) 1993: a verification of some methods to determine the fluxes of momentum, sensible heat, and water vapour using standard deviation and structure parameter of scalar meteorological quantities. *Boundary-Layer Meteorology*, 633(63), 231–257. Available from: https://doi.org/10.1007/BF00710461

De Bruin, H.A.R., Kohsiek, W., Van Den Hurk, B.J.J.M., Van Den Hurk, B.J.J.M. & Kroon, L.J.M. (1999) 1999: on the temperature-humidity correlation and similarity. *Boundary-Layer Meteorology*, 933(93), 453–468. Available from: https://doi.org/10.1023/A:1002071607796

Dyer, A.J. (1974) A review of flux-profile relationships. *Boundary-Layer Meteorology*, 7, 363–372. Available from: https://doi.org/10.1007/BF00240838

Evans, J.G. & de Bruin, H.A.R. (2011) The effective height of a two-wavelength Scintillometer system. *Boundary-Layer Meteorology*, 141, 165–177. Available from: https://doi.org/10.1007/S10546-011-9634-0

Evans, J.G., de Bruin, H.A.R., McNeil, D.D., Finch, J.W., Murray, T., Harding, R.J. et al. (2012) Determination of turbulent heat fluxes using a large aperture scintillometer over undulating mixed agricultural terrain. *Agricultural and Forest Meteorology*, 166, 221–233. Available from: https://doi.org/10.1016/J.AGRFORMET.2012.07.010

Fenner, D., Christen, A., Grimmond, S., Meier, F., Morrison, W., Zeeman, M. et al. (2024) urbisphere-Berlin campaign: investigating multi-scale urban impacts on the atmospheric boundary layer. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BMAS-D-23-0030.1

Foken, T. (2008) *Micrometeorology*. Berlin Heidelberg: Springer, pp. 1–306.

Giorgetta, M.A., Brokopf, R., Crueger, T., Esch, M., Fiedler, S., Helmert, J. et al. (2018) ICON-A, the atmosphere component of the ICON earth system model: I. Model description. *Journal of Advances in Modeling Earth Systems*, 10, 1613–1637. Available from: https://doi.org/10.1029/2017MS001242

GOV.UK. (2023) Population Level and Change. https://www.gov.uk/government/statistics/population-statistics-for-rural-england/a-population-level-and-change

Grange, S.K. (2014) Technical note: averaging wind speeds and directions. pp. 12 https://doi.org/10.13140/RG.2.1.3349.2006

Grimmond, C.S.B., Blackett, M., Best, M.J., Baik, J.J., Belcher, S.E., Beringer, J. et al. (2010) 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, C.S.B., Blackett, M., Best, M.J., Barlow, J., Baik, J.J., Belcher, S.E. et al. (2010) The international urban energy balance models comparison project: first results from phase 1. *Journal of Applied Meteorology and Climatology*, 49, 1268–1292. Available from: https://doi.org/10.1175/2010JAMC2354.1

Grimmond, C.S.B. & Cleugh, H.A. (1994) A simple method to determine Obukhov lengths for suburban areas. *Journal of Applied Meteorology*, 33, 435–440. Available from: https://doi.org/10.1175/1520-0450(1994)033<0435:ASMTDO>2.0.CO;2

Grimmond, C.S.B. & Oke, T.R. (1991) An evapotranspiration-interception model for urban areas. *Water Resources Research*, 27, 1739–1755. Available from: https://doi.org/10.1029/91WR00557

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

Grimmond, C.S.B. & Oke, T.R. (2002) Turbulent heat fluxes in urban areas: observations and a local-scale urban meteorological parameterization scheme (LUMPS). *Journal of Applied Meteorology*, 41, 792–810. Available from: https://doi.org/10.1175/1520-0450(2002)041<0792:THFIUA>2.0.CO;2

Grimmond, C.S.B., Oke, T.R. & Steyn, D.G. (1986) Urban Water Balance 1. *Water Resources Research*, 22, 1397–1403. https://doi.org/10.1029/WR022i010p01397

Grimmond, S. & Ward, H.C. (2021) Urban measurements and their interpretation. In: *Springer handbook of atmospheric measurements*. Cham: Springer, pp. 1391–1423.

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, 1771–1790. Available from: https://doi.org/10.1002/qj.4669

Harman, I.N. & Belcher, S.E. (2006) The surface energy balance and boundary layer over urban street canyons. *Quarterly Journal of the Royal Meteorological Society*, 132, 2749–2768. Available from: https://doi.org/10.1256/qj.05.185

Harman, I.N., Barlow, J.F. & Belcher, S.E. (2004a) Scalar fluxes from urban street canyons part II: model. *Boundary-Layer Meteorology*, 113, 387–410. Available from: https://doi.org/10.1007/s10546-004-6205-7

Harman, I.N., Barlow J.F. Best, M.J. & Belcher, S.E. (2004b) Radiative exchange in an urban street canyon. *Boundary-Layer Meteorology*, 110, 301–316. Available from: https://doi.org/10.1023/A:1026029822517

Hartogensis, O.K., De Bruin, H.A.R. & Van de Wiel, B.J.H. (2002) Displaced-beam small aperture scintillometer test. Part II: CASES-99 stable boundary-layer experiment. *Boundary-Layer Meteorology*, 105, 149–176. Available from: https://doi.org/10.1023/A:1019620515781

Hartogensis, O.K., De Bruin, H.A.R., Van de Wiel, B.J.H., Watts, C.J., Rodriguez, J.C. & De Bruin, H.A.R. (2003) Derivation of an effective height for scintillometers: La Poza experiment in Northwest Mexico. *Journal of Hydrometeorology*, 4, 915–928. Available from: https://doi.org/10.1175/1525-7541(2003)004<0915:DOAEHF>2.0.CO;2

Haugen, D.A. (1973) 1972 Workshop on micrometerology. *Bulletin of the American Meteorological Society*, 54, 108–110.

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, 701–728. Available from: https://doi.org/10.1007/s00704-020-03294-1

Hill, R.J., Lawrence, R.S. & Clifford, S.F. (1980) Refractive-index and absorption fluctuations in the infrared caused by temperature, humidity, and pressure fluctuations. *JOSA*, 70, 1192–1205. Available from: https://doi.org/10.1364/JOSA.70.001192

Kanda, M., Inagaki, A., Miyamoto, T., Gryschka, M. & Raasch, S. (2013) A new aerodynamic parametrization for real urban surfaces. *Boundary-Layer Meteorology*, 148, 357–377. Available from: https://doi.org/10.1007/s10546-013-9818-x

Karsisto, P., Fortelius, C., Demuzere, M., Grimmond, C.S.B., Oleson, K.W., Kouznetsov, R. et al. (2016) Seasonal surface urban energy balance and wintertime stability simulated using three land-surface models in the high-latitude city Helsinki. *Quarterly Journal of the Royal Meteorological Society*, 142, 401–417. https://doi.org/10.1002/QJ.2659

Katul, G., Goltz, S.M., Hsieh, C.-I., Cheng, Y., Mowry, F. & Sigmon, J. (1995) 1995: estimation of surface heat and momentum fluxes using the flux-variance method above uniform and non-uniform terrain. *Boundary-Layer Meteorology*, 743(74), 237–260. https://doi.org/10.1007/BF00712120

Katul, G.G., Finnigan, J.J., Poggi, D., Leuning, R. & Belcher, S.E. (2006) 2006: the influence of hilly terrain on canopy-atmosphere carbon dioxide exchange. *Boundary-Layer Meteorology*, 1181(118), 189–216. https://doi.org/10.1007/S10546-005-6436-2

Kendon, E.J., Roberts, N.M., Fowler, H.J., Roberts, M.J., Chan, S.C. & Senior, C.A. (2014) 2014: heavier summer downpours with climate change revealed by weather forecast resolution model. *Nature Climate Change*, 47(4), 570–576. Available from: https://doi.org/10.1038/nclimate2258

Kent, C.W., Grimmond, S., Barlow, J., Gatey, D., Kotthaus, S., Lindberg, F. et al. (2017) Evaluation of urban local-scale aerodynamic parameters: implications for the vertical profile of wind speed and for source areas. *Boundary-Layer Meteorology*, 1642(164), 183–213. Available from: https://doi.org/10.1007/S10546-017-0248-Z

King, R. (2015) *Comparison of UKV with MORUSES and JULES in urban areas*. Exeter, UK: Met Office.

Kleissl, J., Hartogensis, O.K. & Gomez, J.D. (2010) Test of scintillometer saturation correction methods using field experimental data. *Boundary-Layer Meteorology*, 137, 493–507. Available from: https://doi.org/10.1007/s10546-010-9540-x

Kleissl, J., Hong, S.-H. & Hendrickx, J.M.H. (2009) New Mexico scintillometer network: supporting remote sensing and hydrologic and meteorological models. *Bulletin of the American Meteorological Society*, 90, 207–218. Available from: https://doi.org/10.1175/2008BAMS2480.1

Kohsiek, W., Meijninger, W.M.L., Debruin, H.A.R. & Beyrich, F. (2006) Saturation of the large aperture scintillometer. *Boundary-Layer Meteorology*, 121, 111–126. Available from: https://doi.org/10.1007/s10546-005-9031-7

Kooijmans, L.M.J. & Hartogensis, O.K. (2016) Surface-layer similarity functions for dissipation rate and structure parameters of temperature and humidity based on eleven field experiments. *Boundary-Layer Meteorology*, 160, 501–527. Available from: https://doi.org/10.1007/S10546-016-0152-Y/FIGURES/7

Kormann, R. & Meixner, F.X. (2001) An analytical footprint model for non-neutral stratification. *Boundary-Layer Meteorology*, 99, 207–224. Available from: https://doi.org/10.1023/A:1018991015119

Kotthaus, S. & Grimmond, C.S.B. (2014a) Energy exchange in a dense urban environment – part I: temporal variability of long-term observations in central London. *Urban Climate*, 10, 261–280. Available from: https://doi.org/10.1016/j.uclim.2013.10.002

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

Lagouarde, J.P., Irvine, M., Bonnefond, J.M., Grimmond, C.S.B., Long, N., Oke, T.R. et al. (2006) Monitoring the sensible heat flux over urban areas using large aperture scintillometry: case study of Marseille city during the escompte experiment. *Boundary-Layer Meteorology*, 118, 449–476. Available from: https://doi.org/10.1007/s10546-005-9001-0

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, 1674–1689. Available from: https://doi.org/10.1002/qj.3519

Leclerc, M.Y. & Foken, T. (2014) *Footprints in micrometeorology and ecology*. Berlin Heidelberg: Springer-Verlag, pp. 1–239.

Lemonsu, A., Grimmond, C.S.B. & Masson, V. (2004) Modeling the surface energy balance of the core of an old Mediterranean city: Marseille. *Journal of Applied Meteorology*, 43, 312–327. Available from: https://doi.org/10.1175/1520-0450(2004)043<0312:MTSEBO>2.0.CO;2

Leroyer, S., Bélair, S., Husain, S.Z. & Mailhot, J. (2014) Subkilometer numerical weather prediction in an urban coastal area: a case study over the Vancouver metropolitan area. *Journal of Applied Meteorology and Climatology*, 53, 1433–1453. Available from: https://doi.org/10.1175/JAMC-D-13-0202.1

Lindberg, F. & Grimmond, C.S.B. (2011) Nature of vegetation and building morphology characteristics across a city: influence on shadow patterns and mean radiant temperatures in London. *Urban Ecosystem*, 14, 617–634. Available from: https://doi.org/10.1007/s11252-011-0184-5

Lipson, M.J., Grimmond, S., Best, M., Abramowitz, G., Coutts, A., Tapper, N. et al. (2024) Evaluation of 30 urban land surface models in the urban-PLUMBER project: phase 1 results. *Quarterly Journal of the Royal Meteorological Society*, 150, 126–169. Available from: https://doi.org/10.1002/QJ.4589

Macdonald, R.W., Griffiths, R.F. & Hall, D.J. (1998) An improved method for the estimation of surface roughness of obstacle arrays. *Atmospheric Environment*, 32, 1857–1864 http://www.sciencedirect.com/science/article/B6VH3-3T8GSS3-3/2/5828bb57f43e8e828725bf1925b874ac

Maronga, B. (2014) Monin–Obukhov similarity functions for the structure parameters of temperature and humidity in the unstable surface layer: results from high-resolution large-eddy simulations. *Journal of the Atmospheric Sciences*, 71, 716–733. Available from: https://doi.org/10.1175/JAS-D-13-0135.1

Mason, P.J. (1988) The formation of areally-averaged roughness lengths. *Quarterly Journal of the Royal Meteorological*

Society, 114, 399–420. Available from: https://doi.org/10.1002/QJ.49711448007

Masson, V. (2006) Urban surface modeling and the meso-scale impact of cities. *Theoretical and Applied Climatology*, 84, 35–45. Available from: https://doi.org/10.1007/s00704-005-0142-3

Mcllveen, R. (2010) *Fundamentals of weather and climate*, 2nd edition. Oxford, UK: Oxford University Press, p. 661.

Meijninger, W.M.L., Hartogensis, O.K., Kohsiek, W., Hoedjes, J.C.B., Zuurbier, R.M. & De Bruin, H.A.R. (2002) Determination of area-averaged sensible heat fluxes with a large aperture Scintillometer over a heterogeneous surface – Flevoland field experiment. *Boundary-Layer Meteorology*, 1051(105), 37–62. Available from: https://doi.org/10.1023/A:1019647732027

Meijninger, W.M.L., Hartogensis, O.K., Kohsiek, W., Hoedjes, J.C.B., Zuurbier, R.M., De Bruin, H.A.R. et al. (2005) Scintillometer-based turbulent fluxes of sensible and latent heat over a heterogeneous land surface – a contribution to LITFASS-2003. *Boundary-Layer Meteorology*, 1211(121), 89–110. Available from: https://doi.org/10.1007/S10546-005-9022-8

Moene, A.F. & Schüttemeyer, D. (2008) The effect of surface heterogeneity on the temperature-humidity correlation and the relative transport efficiency. *Boundary-Layer Meteorology*, 129, 99–113. Available from: https://doi.org/10.1007/S10546-008-9312-Z

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.UCLIM.2020.100748

Offerle, B., Grimmond, C.S.B., Fortuniak, K. & Pawlak, W. (2006) Intraurban differences of surface energy fluxes in a central European City. *Journal of Applied Meteorology and Climatology*, 45, 125–136. Available from: https://doi.org/10.1175/JAM2319.1

Oke, T.R. (1988) *Boundary layer climates*, 2nd edition. London: Routledge, p. 435.

Oke, T.R., Mills, G., Christen, A. & Voogt, J.A. (2017) *Urban Climates*. Cambridge, UK: Cambridge University Press.

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, 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, 1530–1542. Available from: https://doi.org/10.1002/qj.572

Reinert, D., Prill, F., Frank, H., Denhard, M., Baldauf, M., Schraff, C. et al. (2021) DWD Database Reference for the Global and Regional ICON and ICON-EPS Forecasting System. pp. 129 https://www.dwd.de/DWD/forschung/nwv/fepub/icon&uscore;database&uscore;main.pdf

Ronda, R.J., Steeneveld, G.J., Heusinkveld, B.G., Attema, J.J. & Holtslag, A.A.M. (2017) Urban Finescale forecasting reveals weather conditions with unprecedented detail. *Bulletin of the American Meteorological Society*, 98, 2675–2688. Available from: https://doi.org/10.1175/BAMS-D-16-0297.1

Schmid, H.P. & Lloyd, C.R. (1999) Spatial representativeness and the location bias of flux footprints over inhomogeneous areas. *Agricultural and Forest Meteorology*, 93, 195–209. Available from: https://doi.org/10.1016/S0168-1923(98)00119-1

Simón-Moral, A., Dipankar, A., Roth, M., Sánchez, C., Velasco, E. & Huang, X.Y. (2020) Application of MORUSES single-layer urban canopy model in a tropical city: results from Singapore. *Quarterly Journal of the Royal Meteorological Society*, 146, 576–597. Available from: https://doi.org/10.1002/qj.3694

Steeneveld, G.J., Tolk, L.F., Moene, A.F., Hartogensis, O.K., Peters, W. & Holtslag, A.A.M. (2011) Confronting the WRF and RAMS mesoscale models with innovative observations in The Netherlands: evaluating the boundary layer heat budget. *Journal of Geophysical Research – Atmospheres*, 116, 23114. Available from: https://doi.org/10.1029/2011JD016303

Stewart, I.D. & Oke, T.R. (2012) Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, 93, 1879–1900. Available from: https://doi.org/10.1175/BAMS-D-11-00019.1

Stull, R.B. (1988) In: Stull, R.B. (Ed.) *An introduction to boundary layer meteorology*, 1st edition. Dordrecht: Springer, p. 670. Available from: https://doi.org/10.1007/978-94-009-3027-8

Tang, Y., Lean, H.W. & Bornemann, J. (2013) The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteorological Applications*, 20, 417–426. Available from: https://doi.org/10.1002/met.1300

United Nations. (2019) World Urbanization Prospects – The 2018 Revision. pp. 103. https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf

Vaisala. (2013) Humidity conversion formulas – Calculation formulas for humidity. pp. 16 https://www.vaisala.com/sites/default/files/documents/Humidity&uscore;Conversion&uscore;Formulas&uscore;B210973EN-F.pdf

Van Ulden, A.P. & Holtslag, A.A.M. (1985) Estimation of atmospheric boundary layer parameters for diffusion applications. *Journal of Applied Meteorology and Climatology*, 24, 1196–1207.

Vesala, T., Kljun, N., Rannik, Ü., Rinne, J., Sogachev, A., Markkanen, T. et al. (2008) Flux and concentration footprint modelling: state of the art. *Environmental Pollution*, 152, 653–666. Available from: https://doi.org/10.1016/J.ENVPOL.2007.06.070

Wang, T.I., Ochs, G.R. & Clifford, S.F. (1978) A saturation-resistant optical scintillometer to measure Cn2. *Journal of the Optical Society of America*, 68, 334–338. Available from: https://doi.org/10.1364/JOSA.68.000334

Ward, H. (2013) Energy, water and carbon fluxes in the suburban environment. PhD Thesis, *King's College London*, 1–243.

Ward, H.C. (2017) Scintillometry in urban and complex environments: a review. *Measurement Science and Technology*, 28, 064005. Available from: https://doi.org/10.1088/1361-6501/aa5e85

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, 65–89. Available from: https://doi.org/10.1007/s10546-014-9916-4

Ward, H.C., Evans, J.G. & Grimmond, C.S.B. (2015a) Infrared and millimetre-wave scintillometry in the suburban environment – part 2: large-area sensible and latent heat fluxes. *Atmospheric Measurement Techniques*, 8, 1407–1424. Available from: https://doi.org/10.5194/amt-8-1407-2015

Ward, H.C., Evans, J.G., Grimmond, C.S.B. & Bradford, J. (2015b) Infrared and millimetre-wave scintillometry in the suburban environment-part 1: structure parameters. *Atmospheric*

Measurement Techniques, 8, 1385–1405. Available from: https://doi.org/10.5194/amt-8-1385-2015

Ward, H.C., Evans, J.G., Grimmond, C.S.B., Bradford, J., Kotthaus, S., Järvi, L. et al. (2016) Surface urban energy and water balance scheme (SUEWS): development and evaluation at two UK sites. *Urban Climate*, 18, 1–32. Available from: https://doi.org/10.1016/j.uclim.2016.05.001

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

Wesely, M.L. (1976) The combined effect of temperature and humidity fluctuations on refractive index. *Journal of Applied Meteorology and Climatology*, 15, 43–49.

Wie, J., Hong, S.O., Byon, J.Y., Ha, J.C. & Moon, B.K. (2020) Sensitivity analysis of surface energy budget to albedo parameters in Seoul metropolitan area using the unified model. *Atmosphere (Basel)*, 11, 120. Available from: https://doi.org/10.3390/ATMOS11010120

Wyngaard, J.C., Izumi, Y. & Collins, S.A. (1971) Behavior of the refractive-index-structure parameter near the ground. *Journal of the Optical Society of America*, 61, 1646. Available from: https://doi.org/10.1364/josa.61.001646

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: Saunders, B., Grimmond, S., Hertwig, D., Bohnenstengel, S.I., Lean, H.W. & Morrison, W. (2024) Methodology to evaluate numerical weather predictions using large-aperture scintillometry sensible heat fluxes: Demonstration in London. *Quarterly Journal of the Royal Meteorological Society*, 1–28. Available from: https://doi.org/10.1002/qj.4837

APPENDIX A. CALCULATION OF METEOROLOGICAL VARIABLES TO DERIVE $Q_{H,LAS}$

As meteorological variables are needed at the LAS-z_f beam height (Section 3.1), corrections are needed from a sensor (e.g., automatic weather station) located at height z_m. In this case, the sensor is at 142 m a.s.l. (Section 4.2.1). These data undergo quality control steps (e.g., logical thresholds and wind direction consistency) before being used for the calculation of $Q_{H,LAS}$.

To correct air pressure (P, units: Pa) we use (Ambaum, 2020):

$$P = P_m \exp\left[\frac{-gM(z_f - z_m)}{R^* T}\right],$$

where P_m is the pressure at height m, g is the gravity acceleration (9.81 m s^{-2}), M is the molar mass of air (0.03 kg mol^{-1}), R* is the universal gas constant (8.31 J mol^{-1} K^{-1}) and T is air temperature (K). Similarly, horizontal wind speed (u, m s^{-1}) is corrected (Oke et al., 2017):

$$u = u_m \frac{\ln\left(\frac{z_f - z_d}{z_0}\right) - \Psi_M\left(\frac{z_f}{L}\right) + \Psi_M\left(\frac{z_0}{L}\right)}{\ln\left(\frac{z_m - z_d}{z_0}\right) - \Psi_M\left(\frac{z_m - z_d}{L}\right) + \Psi_M\left(\frac{z_0}{L}\right)},$$

as part of an iterative process (Figure 3, and Table 1 for notation). The height-corrected u is then used in the calculation of L.

To calculate the SA_{LAS} (Section 3.1.1) u_*, L, wind direction ($\theta°$), and the standard deviation of the v component of wind (σ_v) are needed. Each observation sample (e.g., 5 s) is used to calculate the latter (Grange, 2014):

$$v = -|u| \cos(\theta),$$

prior to determining σ_v for the desired flux averaging period (e.g., 1 min).

Absolute humidity (q, kg m^{-3}) is used in eq. F.2 (Figure 3) to calculate A_T using (Vaisala, 2013):

$$q = \frac{e_w}{R_V \times T},$$

where e_w is the water vapour pressure derived from relative humidity (R_q, %) (Mcllveen, 2010):

$$e_w = e_{ws} \times \frac{R_q}{100},$$

and e_{ws} is the water vapour saturation pressure (Vaisala, 2013):

$$e_{ws} = A \times 10^{\left(\frac{s \times T}{T + T_n}\right)}$$

where values of constants A, s and triple point temperature (T_n) are constant for a given temperature range (Table 1 in Vaisala, 2013).

APPENDIX B. SIMPLIFICATION OF THE STRUCTURE PARAMETER OF THE REFRACTIVE INDEX

As air temperature (T), pressure (P), and absolute humidity (q) can all modify the structure parameter of the refractive index (C_n^2, Section 3.1, eq. F.5, Figure 3), other meteorological variables (Section A.1) are also needed. As variations in P occur at temporal scales that are longer than the averaging period, and on spatial scales larger than L_P (Section 4.2.1), they are typically ignored (Hill et al., 1980), so the refractive index structure parameter can be written (Wesely, 1976) as:

$$C_n^2 = \frac{A_T^2}{\overline{T}^2} C_T^2 + \frac{2\,A_T A_q}{\overline{T}\overline{q}} C_{Tq} + \frac{A_q^2}{\overline{q}^2} C_q^2,$$

where A_T and A_q are the structure parameter coefficients for temperature and humidity, respectively, and C_{Tq} is the covariant term.

A near-infrared single-wavelength (~850 nm) LAS system has greater sensitivity to temperature-caused scintillations than humidity; therefore, the relative contribution to C_n^2 from A_T is much larger than from A_q (Wesely, 1976). If the surface is dry (e.g., by removing wet periods, Section 3.1) the Bowen ratio (β) correction term becomes small and can be ignored (Hill et al., 1980), allowing further simplification to:

$$C_n^2 \approx \frac{A_T^2}{T^2} C_T^2 \left(1 + \frac{0.03}{\beta}\right)^2 \approx \frac{A_T^2}{T^2} C_T^2.$$