Connecting physical and socio-economic spaces for multi-scale urban modelling: a dataset for London

Article

Published Version

Open Access

Geoscience Data Journal

WILEY

Geoscience Data Journal RMetS

DATA SERVICES ARTICLE OPEN ACCESS

Connecting Physical and Socio-Economic Spaces for Multi-Scale Urban Modelling: A Dataset for London

Denise Hertwig[1] | Megan McGrory[1] | Matthew Paskin[1] | Yiqing Liu[2] | Samuele Lo Piano[2] | Heidi Llanwarne[1] | Stefán T. Smith[2] | Sue Grimmond[1]

[1]Department of Meteorology, University of Reading, Reading, UK | [2]School of the Built Environment, University of Reading, Reading, UK

Correspondence: Denise Hertwig (d.hertwig@reading.ac.uk)

Received: 21 June 2024 | **Revised:** 19 November 2024 | **Accepted:** 6 January 2025

Funding: This study was supported by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 855005; project *urbisphere*). UKRI: Natural Environment Research Council (NERC) projects APEx (NE/T001887/2) and ASSURE (NE/W002965/1); and Engineering and Physical Sciences Research Council (EPSRC) project CREDS (EP/R035288/1).

Keywords: agent-based modelling | building energy modelling | building typology | human activity | MAPSECC | urban form and function | urban transport

ABSTRACT

Versatile approaches for urban modelling need to simultaneously consider the physical characteristics of a city (urban form) and urban function as a manifestation of economically, socially, and culturally motivated human activities. Exposure and risk assessment studies concerning urban heat or air pollution can greatly benefit from modelling that dynamically connects physical and socio-economic urban spaces and represents humans as active components of the urban system (e.g., agent-based modelling). The spatio-temporal complexity and variability of urban form, function, human behaviour, and micro-climate put high demands on input data of such models. We present a general methodology for creating a suite of data connecting and harmonising available information for high-resolution modelling. This is demonstrated for London, UK. The multi-scale database covers urban neighbourhoods (at 500 m grid-cell resolution), localised microenvironments of activity, buildings, and extends down to the scale of individuals. Data include neighbourhood land-cover fractions that provide boundary conditions for urban land-surface models and building typologies generated by assessing building function, form, and materials (via building age) that are suitable for building energy modelling. Urban populations (residential, workplace) and demographic composition of households in building typologies are derived. Temporal profiles (10 min resolution) of human activities by age cohort, household size, day type, work patterns, and season derived from time-use survey data are mapped to various socio-economic microenvironments, alongside assessments of activity-dependent electrical energy consumption and human metabolic output. A transport database provides available travel options (1 min resolution) between London neighbourhoods by mode, making use of public transport schedules, road networks, and traffic speeds.

Dataset Information

Identifier: https://doi.org/10.5281/zenodo.12190340

Creators: Hertwig, D., McGrory, M., Paskin, M., Liu, Y., Lo Piano, S., Llanwarne, H., Smith, S.T., Grimmond, S.

Title: Multi-scale harmonisation Across Physical and Socio-Economic Characteristics of a City region (MAPSECC): London, UK

Publisher: Zenodo

Publication year: 2024

Resource type: Dataset

Version: 1.0

1 | Introduction

Representation of the urban micro-climate in environmental modelling is recognised as a topic of increasing importance, with applications from exposure and risk assessment (e.g., heat stress, flooding; London Climate Change Partnership, LCCP 2024) to integrated urban climate services (e.g., WMO's Integrated Urban Hydrometeorological, Climate, and Environmental Services, IUS, Grimmond et al. 2020). The recent increase in the level of physical and process complexity captured in urban climate and land-surface modelling accompanied by high spatio-temporal resolution, has raised demands on the level of detail, resolution, and quality of urban input data (e.g., Ching et al. 2019; Demuzere et al. 2020; Masson et al. 2020). New modelling approaches and applications aim for a dynamic and multi-scale representation of feedback mechanisms between urban climate, urban form (e.g., building geometry, urban surfaces, and materials), function (e.g., building use, urban infrastructure and systems), and human behaviour (e.g., Pfafferott et al. 2021; Takane, Nakajima, and Kikegawa 2022). Modelling human presence and activities in different indoor and outdoor settings is essential to studying exposure to urban environmental stressors like air pollution or heat stress (e.g., Smith et al. 2016; Jia and Wang 2021; Gastineau et al. 2023). Agent-based modelling has emerged as a versatile method to connect physical and socio-economic urban spaces (Yang et al. 2018; Lund, Gouripeddi, and Facelli 2020; Hertwig et al. 2025) by representing interactions between atmospheric state and human behaviour, e.g., by dynamic modelling of urban anthropogenic heat emissions (Capel-Timms et al. 2020).

We present a comprehensive database and methodology (Multi-scale harmonisation Across Physical and Socio-Economic Characteristics of a City region, MAPSECC; Table 1, Figure 1) that describes the physical and socio-economic composition of Greater London, UK (MAPSECC: London). The dataset components combine information from primary sources (listed in Appendix A: Table A1) through novel downscaling and aggregation methods. The processing is repeatable across the UK, as it relies to a large degree on data available from government agencies at the national level. The database spans multiple scales of interest, from neighbourhoods (spatial grid with a horizontal resolution of 500 m; Supporting Information (SM), Section SM1), to buildings, to individual people. Time-varying information on occupancy/activity patterns and locally available transport options (public and private) are processed at resolutions of 10 min and 1 min, respectively, with public transport capacities by mode of travel assessed at a resolution of 10 min. Information on human behaviour is analysed for a variety of sub-grid scale spaces of activity (Section 1.1, Figure 1a). Database components include urban land-cover composition (Section 2; Figure 1b), building typologies (Section 3, Figure 1c) considering form, function and materials, socio-demographic and activity profiles of urban populations (Sections 4 and 5; Figure 1d,e), and a transport database that allows derivation of travel routes (road and rail) for public and private modes of transport (Section 6, Figure 1f).

The database addresses data needs of multi-scale agent-based urban modelling systems like DAVE (Dynamic Anthropogenic actiVities and feedback to Emissions; Hertwig et al. 2025), the successor to Capel-Timms et al.'s (2020) model (data examples given in Section 7). However, all data components have wider (standalone) use, for example (Section 8) for urban-scale building energy modelling, urban land-surface modelling as well as location (e.g., microenvironments) and activity-based exposure assessment studies.

In the following we showcase details of the MAPSECC database, including the main derivation methods with underlying assumptions made in processing of the primary data sources. All details of the processing steps and metadata are documented in SM1–8 (see cross-references in Table 1). This includes documentation of the nature, type, and timing of all primary input data used, together with any assumptions made in their handling during the generation of MAPSECC data. To ensure complete traceability, all processing codes (Python 3) are included in the database. Reference years of the input data (Table A1) are part of the metadata description of all database elements given in the Supporting Information and, for geospatial data components (Table 1), directly included as data attributes.

TABLE 1 | Overview of components of MAPSECC: London (gridded data: 500 m × 500 m horizontal resolution). All geospatial data use the Ordnance Survey (OS) National Grid coordinate reference system (OSGB 1936; EPSG: 27700). Input data sources are given in Appendix A: Table A1. All details on data processing given in SM1–8.

Dataset	Time	Space	Scale	Data type	Data format	Section	SM
Processing grid	—	Gridded	Neighbourhood	Geospatial, vector	ESRI shapefile[a]	1.1	SM1
Land cover	—	Gridded	Neighbourhood	Geospatial, vector	ESRI shapefile	2	SM2
Building typologies & population	—	Gridded	Building & neighbourhood	Geospatial, vector; text	ESRI shapefile, JSON[c]	3.1, 4	SM3
Building materials	—	—	Building	Text	CSV[b]	3.2	SM4
Activity profiles	10 min	—	People	Text	CSV	5	SM5
Travel database	1-, 10 min	Gridded	Neighbourhood	Text	JSON, CSV	6.1	SM6
Road lengths	—	Gridded	Neighbourhood	Geospatial, vector	ESRI shapefile	6.1.1	SM7
Spatial attractors	—	Gridded	Neighbourhood	Text	CSV	6.2	SM8

[a]LoC (2024a).
[b]LoC (2024b).
[c]LoC (2024c).

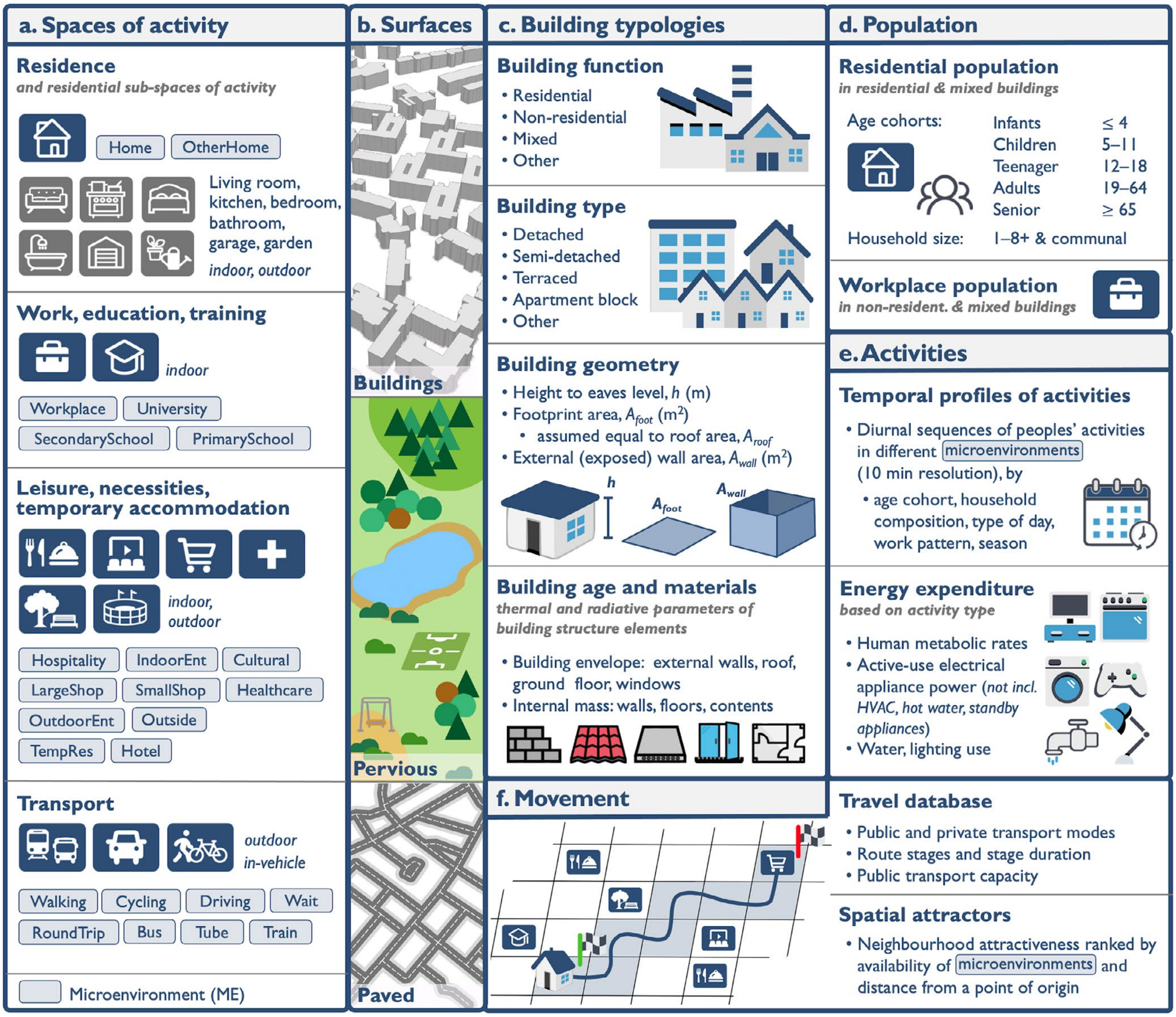

FIGURE 1 | Concepts and components of MAPSECC: (a) places of activity (microenvironments/ME; details in Section 1.1), (b) urban land-surface composition (details Section 2, SM2), (c) building typologies (Section 3, SM3–4), (d) urban populations (Section 4, SM3), (e) human activity profiles and energy expenditure (Section 5, SM5), and (f) transport database to allow movement of people between places of activity (Section 6, SM6–8). Examples of aspects of the database are presented in Section 7.

1.1 | Spaces of Activity

Modelling interactions between humans and the city environment requires defining settings based on locations in space and activities (spaces/environments). While the spatial resolution of locations in the database is constrained by the neighbourhood-scale processing grid (500 m grid length, Section SM1), analysis of the socio-economic context of each grid cell determines the sub-grid scale spaces of activity used in different dataset components (Figure 1a,e,f).

We define four groups of spaces: a citizen's residence; their place of work or education; places for leisure, necessities, or social activities; and transport (Figure 1a). Each of these includes different microenvironments of activity as shown in Table 2a–d. Following environmental exposure assessment terminology (e.g., Duan 1982), we define a microenvironment (ME) as a setting where a person spends parts of their day. These are not necessarily connected to a specific location in space and can also represent a state (Table 2e). Settings may be indoors and/or outdoors (Table 2), and people may carry out very different activities in each ME (e.g., 'Home' *cf.* 'Workplace'). Some of the occupational and non-occupational MEs are based on UK time use survey activity types and locations (TUS; Gershuny and Sullivan 2017; details in Section 5) and are subdivided further considering differences in energy consumption patterns and exposure levels to environmental factors (e.g., to define non-residential building typologies, Section 3.1.5).

2 | Urban Land-Cover Fractions

Urban weather/climate variables like radiation, turbulent heat fluxes, or (near-)surface temperatures can be modelled

TABLE 2 | Microenvironments (ME) considered for people's activities and movement within the city (Sections 5 and 6). (a)–(d) as in Figure 1a, (e) 'Sleep' is the only ME representing a state (as opposed to 'awake'). Residential sub-spaces in (a) are highlighted in italics. MEs with * are not in the activity dataset derived from UK–TUS data (Section 5).

	ME/*sub-spaces*	Context	Description of place or setting
(a) Residences	Home	Indoor/outdoor	Citizen's residence
	OtherHome	Indoor/outdoor	Residence of another citizen
	Kitchen	Indoor	Residential sub-space
	Living room	Indoor	Residential sub-space
	Bedroom	Indoor	Residential sub-space
	Bathroom	Indoor	Residential sub-space
	Garage	Indoor	Residential sub-space
	Garden	Outdoor	Residential sub-space
(b) Work, education, training	Workplace	Indoor	Workplace
	University	Indoor	Higher education (university, college)
	PrimarySchool	Indoor	Primary education
	SecondarySchool	Indoor	Secondary education
(c) Leisure, necessities, temporary accommodation	LargeShop	Indoor	Large scale commercial, retail settings (e.g., chain supermarket, department store, do-it-yourself store)
	SmallShop	Indoor	Small to medium scale commercial, retail or service settings (e.g., convenience store, bank, hairdresser)
	Hospitality	Indoor	Restaurant, café, bar, etc.
	Outside	Outdoor	Public parks, recreational greenspace, playgrounds
	OutdoorEnt	Indoor/outdoor	Outdoor entertainment (e.g., zoo, theme park, sports park)
	IndoorEnt	Indoor	Indoor entertainment with high energy use (e.g., cinema, theatre, ice rink, swimming pool)
	Cultural	Indoor	Indoor cultural/social settings with low to medium energy use (e.g., museum, library, community centre, place of worship)
	Hotel	Indoor	Hotel, hostel, guesthouse, etc.
	Healthcare*	Indoor	Hospital, doctor's surgery, health facility, etc.
	TempRes*	Indoor	Temporary residence (e.g., nursing home, prison)
(d) Transport	Driving*	In-vehicle	Private transport mode (motorised vehicle)
	Cycling*	Outdoor	Private transport mode (bicycle)
	Walking*	Outdoor	Private transport mode (by foot)
	Bus*	In-vehicle	Public transport mode
	Tube/subway*	In-vehicle	Public transport mode
	Train*	In-vehicle	Public transport mode
	RoundTrip	Outdoor	Travel activity starting and ending in the same ME (e.g., walking a dog, accompanying a child)
(e) State	Sleep	Indoor/outdoor	Sleep (awake if in any of the other MEs)

with urbanised land-surface models (e.g., Lipson et al. 2024). Such models need characterisation of the urban surface cover (Figure 1b). We map Geomni UKMap (2021) land-cover data (vector polygons) for Greater London to seven surface types (Table 3a; visualised in Section 7, Figure 4) based on the input data feature type code (FTC) and aggregate data in 500 m grid-cell neighbourhoods.

Vegetated areas with trees in Geomni UKMap (2021) are given in mixed-type categories of scattered trees with a surface cover between 30% and 70% (FTC = 6) and > 70% (FTC = 7). The mid-points of these ranges are used to split the FTC = 6 and 7 areas into tree and non-tree fractions, using $\varphi_{\mathrm{FTC},6} = 0.5$ and $\varphi_{\mathrm{FTC},7} = 0.85$ as factors. The remaining non-tree area fraction is assigned to bare soil ($\varphi_{\mathrm{soil}} = 0.2$) and grass ($\varphi_{\mathrm{grass}} = 0.8$), based on values proposed for London by Ward and Grimmond (2017). Therefore, the total plan area of grass (A_{grass}) is

$$\begin{aligned} A_{\mathrm{grass}} = {} & A_{\mathrm{FTC},3} + A_{\mathrm{FTC},6}\left(1-\varphi_{\mathrm{FTC},6}\right)\varphi_{\mathrm{grass}} \\ & + A_{\mathrm{FTC},7}\left(1-\varphi_{\mathrm{FTC},7}\right)\varphi_{\mathrm{grass}}\ \left(\mathrm{m}^2\right) \end{aligned} \quad (1)$$

and the total plan area of bare soil (A_{soil}) is

$$\begin{aligned} A_{\mathrm{soil}} = {} & A_{\mathrm{FTC},8} + A_{\mathrm{FTC},9} + A_{\mathrm{FTC},6}\left(1-\varphi_{\mathrm{FTC},6}\right)\varphi_{\mathrm{soil}} \\ & + A_{\mathrm{FTC},7}\left(1-\varphi_{\mathrm{FTC},7}\right)\varphi_{\mathrm{soil}}\ \left(\mathrm{m}^2\right) \end{aligned} \quad (2)$$

TABLE 3 | Geomni UKMap (2021) feature type codes (FTC) and feature classification codes (FCC1) are used to derive (a) land cover and (b) split of greenspace between residential and recreational.

	UKMap FTC	Processing
(a) Land-cover type		
Building	1, 5	/
Paved (impervious)	2	/
Water	4	/
Grass	3, 6, 7	Equation (1)
Bare soil	8, 9, 6, 7	Equation (2)
Deciduous trees	6, 7	Equation (3)
Evergreen trees	6, 7	Equation (4)
(b) Other categories		
Residential (private) greenspace	3, 6, 7, 8, 9	Equation (5) with FCC1 = 19
Recreational greenspace	3, 6, 7, 8, 9	Equation (5) with FCC1 = 15
All other greenspace	3, 6, 7, 8, 9	Equation (5) with FCC1 ≠ 15,19

Note: UKMap FTC categories: 1: building, 2: man-made [*sic*] surfaced area, 3: vegetated/natural (predominantly grass and low vegetation), 4: water, 5: man-made [*sic*] structure other than building, 6: vegetated/natural area (scattered trees, > 30% and < 70% cover), 7: vegetated/natural area – predominantly trees (> 70% cover), 8: not used, 9: non-geographic entity type. UKMap FCC1 categories: 15: recreation and leisure, 19: residences.

The tree plan area is split between deciduous and evergreen trees using $\varphi_{\mathrm{dectree}} = 0.8$ and $\varphi_{\mathrm{evetree}} = 0.2$ based on London values from Ward and Grimmond (2017):

$$A_{\mathrm{dectree}} = A_{\mathrm{FTC},6}\ \varphi_{\mathrm{FTC},6}\varphi_{\mathrm{dectree}} + A_{\mathrm{FTC},7}\ \varphi_{\mathrm{FTC},7}\varphi_{\mathrm{dectree}}\ \left(\mathrm{m}^2\right) \quad (3)$$

$$A_{\mathrm{evetree}} = A_{\mathrm{FTC},6}\ \varphi_{\mathrm{FTC},6}\varphi_{\mathrm{evetree}} + A_{\mathrm{FTC},7}\ \varphi_{\mathrm{FTC},7}\varphi_{\mathrm{evetree}}\ \left(\mathrm{m}^2\right) \quad (4)$$

Whilst the Geomni UKMap (2021) polygons have very high spatial resolution, individual trees, such as scattered along streets, or other small-scale roadside greenery co-occurring with predominantly paved surfaces are not identified.

The total greenspace (A_{green}, m^2, including bare soil) given by

$$\begin{aligned} A_{\mathrm{green}} &= A_{\mathrm{FTC},3} + A_{\mathrm{FTC},6} + A_{\mathrm{FTC},7} + A_{\mathrm{FTC},8} + A_{\mathrm{FTC},9} \\ &= A_{\mathrm{grass}} + A_{\mathrm{dectree}} + A_{\mathrm{evetree}} + A_{\mathrm{soil}} \end{aligned} \quad (5)$$

is further categorised using the Geomni UKMap (2021) Level 1 feature classification codes (FCC1, Table 3b) including recreational greenspace, A_{grec}, (FCC1 = 15; recreation and leisure), residential (private) greenspace, A_{gres}, (FCC1 = 19; residences) as well as all remaining (other) types of greenspace, A_{goth}, such that $A_{\mathrm{green}} = A_{\mathrm{grec}} + A_{\mathrm{gres}} + A_{\mathrm{goth}}$.

Land-cover fractions (f_i – dimensionless) by surface type i (Table 3) for each neighbourhood grid cell are derived by dividing their respective plan area falling within the grid boundaries, A_i (m^2), by the grid-cell total area of all types, $A_{\mathrm{total}} = \sum_i A_i$ (m^2), so that for each grid box $\sum_i f_i = 1$.

3 | Building Typologies and Materials

3.1 | Building Typologies

Derivation of neighbourhood (grid cell) information characterising building typologies (Figure 1c) is based on downscaling and mapping of physical and socio-economic data (Table A1) onto OS MasterMap (2019) building footprints (all technical details presented in SM3). We use the 2011 English census regions (ONS 2011a) and 33 London borough identifiers, including the City of London (London Datastore 2014), to geolocate each OS MasterMap (2019) building polygon in the process.

Table 4 gives an overview of information derived from, and mapped to, individual OS MasterMap (2019) buildings. The local data-enriched building stock information is aggregated at the 500 m grid scale into building typologies and morphology statistics (Section 3.1.5, Table 5). The former can be used as representative building archetypes in city-scale building energy modelling while the latter provides consistent morphology properties for urban land-surface models.

3.1.1 | Building Morphology

Building morphology parameters (Table 4c) are calculated from OS MasterMap (2019) vector polygons, which already include

TABLE 4 | Overview of (a–e) information derived from or downscaled to individual OS MasterMap (2019) building polygons with height attributes. Building materials (f) can be linked to building typologies via the building age band (e). Adjustments of the building height, h_b (c), in case of implausible input data values are discussed in Section SM3.

	Options and parameters			Section
(a) Function	• Residential • Non-residential • Mixed (residential and non-residential) • Other (uninhabitable, e.g., garage, shed, outbuildings)			3.1.2
(b) Type	• Detached (including bungalows) • Semidetached • Terraced • Apartment blocks (purpose-built flats) • Other (none of the above) • Shed (if uninhabitable structure)			3.1.3
(c) Geometry	• Main buildings (*parents*) and building elements (*daughters*) identified • Roof geometry: flat with $A_{\mathrm{roof}} = A_{\mathrm{foot}}$			3.1.1
	Parameter	**Notation**	**From**	
	Height to eaves level (m)	h	OS input data	
	Maximum height (m)	h_{max}	OS input data	
	Height (m)	h_{b}	h (details Section SM3)	
	Footprint area (m^2)	A_{foot}	GIS derived	
	External wall length (m)	L	GIS derived	
	External wall area (m^2)	A_{wall}	$A_{\mathrm{wall}}=L\cdot h_{,b}$	
	Volume (m^3)	V	$V=A_{\mathrm{foot}}\cdot h_b$	

(d) Property	• Residential property numbers by type (1–4+ bedrooms)
	  ◦ Detached, semi-detached, terraced
	  ◦ Flats/apartments (purpose-built or converted)
	• Used in downscaling of residential populations (Section 4)
				3.1.4

(e) Age band	• Residential, mixed: 15 periods from pre–1900 to 2018–2020 • Non-residential: 6 periods from pre–1940 to 2001–2003			3.1.4
(f) Materials	• Mappable from building age band information (e), for ◦ External walls, roofs, ground floors, windows, internal walls, internal floors			3.2
	Parameter		**Notation**	
	Effective thickness (m)		d_{eff}	
	Effective thermal conductivity (W m^{-1} K^{-1})		k_{eff}	
	Effective density (kg m^{-3})		ρ_{eff}	
	Effective specific heat capacity (J kg^{-1} K^{-1})		$c_{\mathrm{p,eff}}$	
	Thermal mass (external walls, ground floors, roofs only) (J m^{-2} K^{-1})		C_{eff}	
	Emissivity of the external surface		$\varepsilon_{\mathrm{ex}}$	
	Emissivity of the internal surface		$\varepsilon_{\mathrm{in}}$	
	Absorptivity of the external surface		φ	
	Reflectivity (or albedo) of the external surface		α	
	Transmissivity of the external surface		τ	

attributes of eaves height, h (m), and maximum height above ground level. The building height, h_b, used in further calculations of geometry attributes is equal to h, but adjusted when rare invalid (e.g., NULL or 0 m) values occur in the input data (details in Section SM3). Calculated for each building polygon are: footprint area, A_{foot} (m^2), volume, $V = A_{foot} \bullet h_b$ (m^3), length of external walls (i.e., non-party walls; in contact with the outside air), L_{wall} (m), and external wall area, $A_{wall} = L_{wall} \bullet h_b$ (m^2). Since heights are given only up to eaves level and no further information on local roof types is available from the input data, by default all roofs are assumed to be flat with a roof area of $A_{roof} = A_{foot}$.

In the OS MasterMap (2019) dataset, an individual building may consist of multiple polygons. For example, large height differences may exist between building elements, such as annexes and balconies, or because of roof structure complexity. Similarly, garages, sheds, or other small-scale uninhabitable buildings may exist as attached or standalone structures but are excluded in some subsequent processing steps (e.g., population downscaling; Section 4). Hereafter, we refer to the main building as a *parent* and any building elements or smaller uninhabitable structures as *daughters*. Daughter buildings are identified based on their size and shape (Section SM3.3). Parent–daughter pairs that share a wall do not have overlapping building footprints. Hence, to derive morphology parameters for the entire building, parent–daughter groups are identified in an iterative process. Geometry parameters of the complete building, accounting for all its polygon components, are created and used in the derivation of building typologies (Section SM3, Table SM3.6).

3.1.2 | Building Function

Four building function categories are distinguished: (i) residential, (ii) non-residential, (iii) mixed (shared residential and non-residential use), and (iv) other (i.e., uninhabitable structures like garages, sheds, or outbuildings identified as standalone daughter buildings; Section 3.1.1). Ordnance Survey information on the location and nature of key functional and infrastructure sites (OS MasterMap 2022) and of private and public businesses and services (OS POIS 2023) is downscaled to building polygons. Further mapping of land-use and points of interest from OpenStreetMap crowd-sourced data (OSM 2020) provides complementary information (e.g., industrial, military and outdoor leisure sites).

Given the lack of detailed information, the potential for buildings to be mixed-use is assessed from the type of non-residential building function and land-use information. Excluding settings like hospitals, primary to tertiary education, stadiums, police stations, zoos, military barracks, and industrial parks from residential use, mixed-use is allowed for settings of hospitality, commercial/retail, and service sectors. The downscaling of household populations considers residential and mixed-use buildings, with priority given to the former (Section 4).

3.1.3 | Building Type

Each parent building is assigned a descriptive building type based on its morphology and the number of parent neighbour buildings with which a wall is shared. The following building types are used across all building function categories:

- detached (including bungalows)
- semi-detached
- terraced
- apartment block
- other (none of the previous categories applies).

Daughter buildings of building function 'other' (uninhabitable structures, Section 3.1.2) are labelled as 'shed.'

3.1.4 | Residential Property Type and Building Age Band

To guide the downscaling of residential populations and household sizes (Section 4.1), the number of residential properties by type is mapped to residential and mixed-use parent buildings from building stock information (VOA 2020a) available in the 2011 census Lower layer Super Output Areas (LSOA; ONS 2011a). Input data give property numbers per type (bungalow, terraced, semi-detached, detached, and flats/apartments) and number of bedrooms (1 to 4+) rounded to the closest multiple of 5 per LSOA. These are paired with OS MasterMap (2019) polygons by considering building location, type (Section 4.1.3), and size (i.e., larger properties are mapped to larger buildings). Flat properties are first assigned to apartment blocks (purpose-built flats). Remaining flats are mapped to other inhabitable building types, hence covering instances of dwelling conversion (e.g., from single- to multi-dwelling terraced houses).

Building age bands can be used to assess built-period typical construction types and associated thermal and radiative parameters required in building energy modelling (Section 3.2). LSOA-level residential building age bands (15 periods from pre–1900 to 2018–2020; VOA 2020b) are downscaled to residential and mixed-use parent buildings by creating a link to the VOA (2020a) property types via their distribution across council tax bands. Age bands of commercial and industrial buildings are downscaled from information at the London borough level (six periods from pre–1940 to 2001–2003; MHCLG 2004) by sector (retail, offices, warehouses, factories).

3.1.5 | Neighbourhood-Scale Aggregation

Building-level information (Table 4a–e) is aggregated for each 500 m grid-cell neighbourhood into local building typologies (archetypes) for building energy modelling and grid-level statistics (e.g., mean heights needed in urban land-surface modelling). Table 5 describes dataset components with details given in Section SM3.

3.2 | Building Materials

With age bands mapped to buildings (Section 3.1.4), period-specific construction types and building materials can be

TABLE 5 | Overview of database components of (a) building typologies and (b) grid-level building statistics derived from data-enriched OS MasterMap (2019) buildings (Table 4) in 500 m grid cells. All morphology parameters are derived from combined buildings (parent–daughter pairs; Section 3.1.1). Examples of data are given in Section 7, Figures 5c,d and 6.

(a) Building typologies

Segmentation of typology instances based on combination of
- Function (residential, non-residential, mixed-use, other)
- Type (detached, semi-detached, terraced, apartment block, other, shed)
- Non-residential function type (based on some MEs, Table 2; details listed in Table SM3.8)

Some typology attributes are listed below (all details in Tables SM3.17–SM3.20)

Parameter	Info
Building count	Number of buildings in typology instance
Property type count	Number of buildings with a single or multiple properties/households
Occupant count	Totals and in buildings with a single or multiple properties/households
Household size count	Household numbers by sizes 1–8+ and communal (Section 4.1, Table 6a)
Age cohort count	Occupant numbers by five age groups (Section 4.1, Table 6a)
Age band	Most frequent occurrence in typology instance
Height (h_b, m)	Footprint area-weighted mean, median, 25th, 75th percentiles of buildings in instance
Footprint area (A_{foot}, m^2)	Mean, median, 25th, 75th percentiles of buildings in instance
External wall area (A_{wall}, m^2)	Mean, median, 25th, 75th percentiles of buildings in instance

(b) Grid-level building statistics

Parameter	Info
Total building volumes (V, m^3)	Grid total by building function, building type (residential, mixed-use only)
Building type fractions	Grid ratio of building type volumes to total volume (residential, mixed only)
Height (h_b, m)	Grid mean by building function
Standard deviation of height (e_{h_b}, m)	Grid standard deviation by building function
Maximum height ($h_{b,max}$, m)	Grid maximum by building function

evaluated for each building type and neighbourhood. This is needed for building-level modelling applications that require information on thickness and thermal and radiative parameters of building elements to model temperatures (e.g., of surfaces, indoor air) and heat exchange with the environment. Effective parameters are derived that are representative of the overall building element layer composition (Table 4f) for external walls, roofs, windows, ground floors, internal walls, and internal floors/ceiling. This uses reference typologies that reflect national standards (i.e., country-wide in the UK context), so no regional variations specific to London buildings are considered. Where regional standards exist, these can be implemented (e.g., USA: ICC 2021, China: MHURD 2015).

The energy performance of building envelope components by age band for different UK residential building typologies is described in TABULA (2016) via the thermal transmittance or U-value (Loga et al. 2016). The U-value (W m^{-2} K^{-1}) measures the efficiency with which heat is transferred through the fabric of the building. A low U-value indicates low heat loss (i.e., the building fabric insulates well). In a process that conserves the U-value of the TABULA (2016) 'as-built' typologies, we estimate the structure layer composition using reference types and material parameters from CIBSE (2015) engineering guidelines for external walls, roofs, and ground floors. For windows, U-values and solar heat gain coefficients (*SHGC* or g-value) from TABULA (2016) are used to derive effective thermal and radiative parameters using the EnergyPlus simple window model (DOE 2024; their Section 7.7.4). Since no reference data are available from TABULA (2016) for internal building components (walls, floors/ceilings), CIBSE (2015) reference structures are used. Effective thermal and radiative parameters (Table 4f) by built-period type are derived for all six structural components and can be linked to building typologies (Section 3.1.5) to provide input parameters for urban building energy models. Detailed processing steps are given in Section SM4.

4 | Urban Populations

Residential (households and communal living; Section 4.1) and workplace populations (Section 4.2) are downscaled to individual OS MasterMap (2019) parent buildings by considering

TABLE 6 | Overview of (a) residential and (b) workplace populations mapped to OS MasterMap (2019) buildings and description of presentation within the dataset. Age cohort descriptors in (a) reflect assumptions on primary influencing factors on location and timing of activities, while intra-cohort variations (e.g., economic (in-)activity) are accounted for in the TUS-based activity profiles (Section 5).

Population	Parameters	Section	Dataset components
(a) Residential	• Household size (1–8+), communal • Age cohort aggregation of the 16 census age groups: **Cohort** / **Age** / **Description** Infant / ≤4 / Fully dependent Child / 5–11 / 1° school age Teenager / 12–18 / 2° school age Adult / 19–64 / Working age Senior / ≥65 / Retirement age	4.1	• Population totals (all residents and by age cohort) in 500 m grid cells (Figure 5a) • Household size counts in 500 m grid • Population totals by age cohort and counts of household sizes (including communal living) in building typologies (Section 3.1.5)
(b) Workplace	• Number of workers per non-residential or mixed-use building	4.2	• Totals in 500 m grid cells (Figure 5b)

the building characteristics assigned (Section 3.1). Parameters mapped at building scale are shown in Table 6, together with their representation in the database (processing details in Section SM3).

4.1 | Residential Population

The distribution of residential household populations (household sizes: 1–8+ persons) is available at Output Area (OA) scale from the 2011 census data (ONS 2011b). Downscaling to residential and mixed-use parent buildings per OA uses mapped property numbers by type (Section 3.1.4). In an iterative process, first the largest households are assigned to single-property buildings with the most bedrooms (4+), with remaining households assigned to converted dwellings and apartment blocks. This assumes flats, on average, have smaller household sizes than single-family houses. Communal populations (i.e., residents in shared accommodations/communal living) from ONS (2011b) are mapped to non-residential buildings with educational (e.g., halls of residence) and military (barracks) settings if available and to residential buildings otherwise. The number of communal residents assigned is based on building volume, assuming a floor area of 15 m^2 per person when living in shared accommodation (typical unit size in purpose-built student accommodations in London).

Household and communal populations have age distributions across 16 census age groups (0–4 to 90+; ONS 2011b) and are aggregated to five cohorts (Table 6a) considering primary influencing factors on location and timing of activities (dependency, school, work, retirement; Section 5). As details on the demographic composition of households are not available from the census data, age distributions are assigned to households within a building using a randomised process that considers household size and ensures children under the age of 18 live with at least one adult. Communal living settings are assumed to predominantly have an age composition typical for each setting (e.g., university/college).

4.2 | Workplace Population

OA-level workplace populations ($\mathrm{WPop}_{\mathrm{tot}}^{\mathrm{OA}}$; ONS 2011c) are mapped to non-residential and mixed-use parent buildings. The number of workers ($\mathrm{WPop}_{i}^{\mathrm{OA}}$) per individual building, i, in an OA is based on the ratio of the building's volume (V_i^{OA}) to the combined total building volume ($V_{\mathrm{tot}}^{\mathrm{OA}} = \sum_{i=1}^{N} V_i^{\mathrm{OA}}$) of all N non-residential and mixed-use buildings using an iterative process:

$$\mathrm{WPop}_i^{\mathrm{OA}} = \begin{cases} \left[\dfrac{V_i^{\mathrm{OA}}}{V_{\mathrm{tot}}^{\mathrm{OA}}} \mathrm{WPop}_{\mathrm{tot}}^{\mathrm{OA}} \right] & \text{if } i = 1 \\ \left[\dfrac{V_i^{\mathrm{OA}}}{\left(V_{\mathrm{tot}}^{\mathrm{OA}} - \sum_{j=1}^{i-1} V_j^{\mathrm{OA}}\right)} \left(\mathrm{WPop}_{\mathrm{tot}}^{\mathrm{OA}} - \sum_{j=1}^{i-1} \mathrm{WPop}_j^{\mathrm{OA}}\right) \right] & \text{if } i > 1 \end{cases} \tag{6}$$

with $i = 1, \ldots, N$ and $[\ldots]$ denotes rounding to the nearest integer. For mixed-use buildings, only half of the volume is considered as non-residential in the downscaling. The second condition ($i > 1$) in Equation (6) ensures that rounding does not result in differences between the input data and mapped total workplace population.

5 | Human Activities and Energy Expenditure

Modelling anthropogenic heat emissions or human exposure requires information on type, duration, and location of activities and recognition of the impact of socio-economic factors (e.g., type of day, age cohort, work pattern, household characteristics; Figure 1e). Time series of people's activities (hereafter: 'activity profiles', Section 5.1) are generated from UK time use survey (TUS) information (UK–TUS 2014–15; Gershuny and Sullivan 2017), reflecting transitions between and duration of stay within microenvironments (Table 2; examples in Section 7, Figure 7a–c). Agent-based models can use such data to initiate the movement of modelled citizens between MEs in time and space (Section 6). The activity profiles include metabolic and residential appliance energy expenditure by activity type (Section 5.2). This allows building energy models to capture the dependency of internal heat gains on occupancy patterns, household composition, size, and other factors (e.g., Liu et al. 2024; Hertwig et al. 2025). All details of the processing are given in Section SM5.

5.1 | Human Activity Profiles

UK–TUS 2014–15 data (Gershuny and Sullivan 2017) are based on self-completed diaries and household interviews conducted between April 2014 and December 2015. The surveys completed across the UK occur relatively uniformly across seasons. However, data becomes sparser once further stratified (Appendix B: Table B1). For each diarist (age 8 and above), details on activity and setting are surveyed using 10 min timesteps over a 24-h period (4 am to 4 am local time, i.e., 144 timesteps over 24 h) on both a weekday and weekend day. The diarists' socio-demographic background (e.g., age, household size) and work pattern are captured. For each timestep, a primary activity plus up to three further co-occurring activities are documented. Across the entire survey, < 17%, < 2%, and < 0.2% of timesteps have secondary, tertiary, and quaternary activities listed, respectively. UK–TUS activity types can be grouped into categories of self-care (e.g., sleeping, personal hygiene), care work (e.g., childcare, help to other households), chores (e.g., housekeeping, shopping, repairs), occupational activities (work, education), leisure, and physical exercise/sports (Figure 2).

From each TUS diary, a diurnal sequence of activity profiles is generated with a new profile being added after a change of ME or state (sleeping, awake) occurs (see example sequences in Table B2). Profiles give the start timestep and the full list of primary activities carried out in the current ME at each timestep before a transition to the next ME or state.

Each activity profile includes metadata on (i) age cohort (Table 6a), (ii) household size, (iii) day type (weekend, weekday with school, weekday without school), (iv) work pattern (workday/school day, non-work/non-school day), and (v) season the activity occurred (spring–MAM, summer–JJA, autumn–SON, winter–DJF, assuming 3 months per season as indicated by their first letter). School holiday dates are derived for the TUS years from the Royal Borough of Kensington and Chelsea

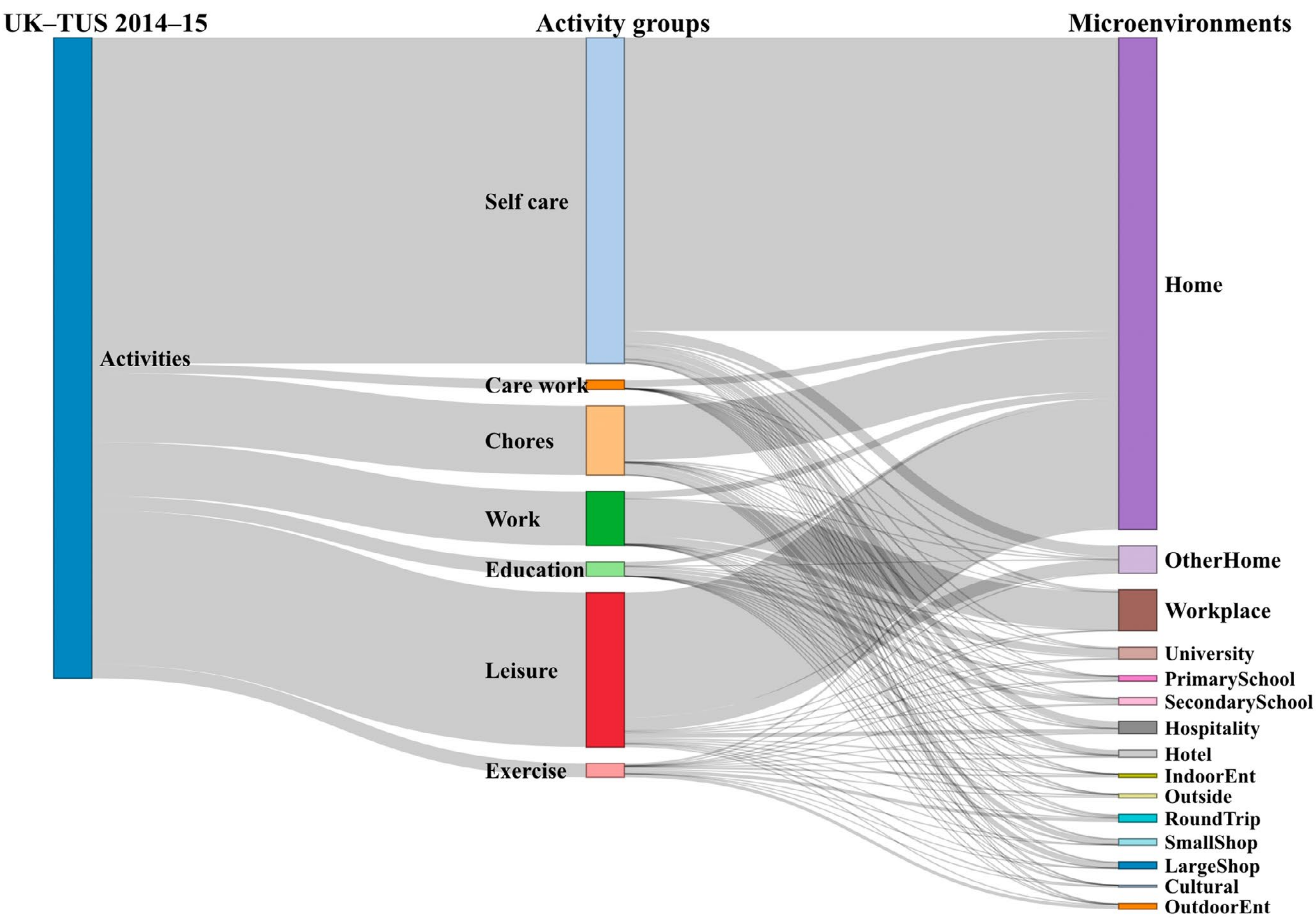

FIGURE 2 | Mapping of UK TUS 2014–15 (Gershuny and Sullivan 2017) activities to activity groups and microenvironments (Table 2). Proportions of splits are based on the total number of 10-min surveyed timesteps across the complete dataset (all age groups, household sizes, day types, seasons). Timesteps in the 'Sleep' state ME are assigned to the MEs the diarist fell asleep in (e.g., 'Home').

(RBKC 2014, 2015, 2016), chosen due to public data availability during the TUS survey period. Regional deviations of UK school holidays are not accounted for. The youngest UK–TUS diarists are 8 years old; however, we extrapolate the child cohort to 5-year-olds as it is assumed the daily schedules in the 5–8 age bracket are similar to that of children of age 8–11, who all typically attend primary school (Table 6a; age 5–11). Infants' time allocation (age 0–4) is not surveyed, but for modelling purposes this group can be treated as dependents (i.e., infants do not travel unaccompanied and do not have energy-use profiles separate from their caretakers).

5.1.1 | Mapping of Microenvironments

Activity profiles are created for individual members of the four non-infant age cohorts (Table 6a) by associating the primary UK–TUS activity type and location with a microenvironment (Table SM5.5), allowing the timing of transitions between and durations of stay within MEs at the 10-min survey timestep. If multiple MEs are associated with a UK–TUS location, further factors such as activity type and age group are considered. For example, for surveyed children (age 8–11), the UK–TUS location 'Working place or school' is mapped to the 'PrimarySchool' ME, while the type of activity specified by the diarists determines whether adults or seniors are at their 'Workplace' or 'University', or if teenagers (age 12–18) are in 'SecondarySchool' or 'Workplace.' In the case of other ambiguous instances (e.g., mapping of the TUS location 'Shopping centres, markets, other shops' to either 'LargeShop' or 'SmallShop'), the mapping is based on an equal-weighted random choice. If no location was specified by the diarist, look-up tables are used to probabilistically associate activities with their most likely ME (details in Section SM5.3).

The time taken to travel between MEs is included in the UK–TUS data as travel activity. In the mapping, the travel timesteps reported are added onto the following activity profile. For example, for the change between 'Home' and 'Workplace', the travelling (commuting to work) is part of the 'Workplace' activity profile as in the example shown in Table B2a. Transport MEs for travel between two different MEs can be allocated dynamically in the modelling, e.g., based on route and mode options available at the time of travel (Section 6). Short trips that start and end in the same ME (e.g., walking the dog) are mapped directly from UK–TUS into a separate 'RoundTrip' ME (Table 2).

We identify one microenvironment, 'Sleep' (Table 2), from the corresponding TUS activity that is a state rather than a setting and differs from all other activities (i.e., awake). The citizen ME setting while in the 'Sleep' state is the ME they fell asleep in (e.g., 'Home', 'Hotel', or other as derived from the diarist's response).

For the residential 'Home' and 'OtherHome' MEs, at each timestep, activities are further associated with the sub-space (or 'nanoenvironment') in which they most likely occur, i.e., kitchen, living room, bedroom, bathroom, garage, or garden (Table 2; Figure 7d). This enables the monitoring of the occupancy of indoor and outdoor residential spaces.

5.2 | Energy Expenditure

5.2.1 | Metabolic Rates

Human metabolic rates for UK–TUS primary activities are derived at each 10-min timestep by mapping the TUS activities to classifications of energy expenditure by activity type identified by Ainsworth et al. (1993) and ASHRAE (2017). These are available in terms of dimensionless MET units that characterise the intensity of a specific activity, $\mathrm{MET}_{\mathrm{act}}$, as the ratio of the metabolic rate for that activity (M_{act}) and the resting (or basal) metabolic rate (sitting quietly, M_{rest}). If no direct equivalent to a specific TUS activity is available, the most similar activity in terms of overall intensity is used. The metabolic rate for each activity, M_{act}, is calculated as

$$M_{\mathrm{act}} = \mathrm{MET}_{\mathrm{act}}\, M_{\mathrm{rest}}\, A_{\mathrm{body}}\, f_{\mathrm{age}} \ (\mathrm{W}) \tag{7}$$

where the resting metabolic rate M_{rest} is set to $55\,\mathrm{W\,m^{-2}}$ (BS EN ISO 8996 2004). M_{act} in Equation (7) considers age-dependence via A_{body} and f_{age}, but no further factors (e.g., sex/hormonal), and input data used represent averages over male and female cohorts. The body surface area, A_{body} (m^2), is computed following Mosteller (1987) as $A_{\mathrm{body}} = \sqrt{(h_{\mathrm{body}} \cdot m_{\mathrm{body}})/c}$ where h_{body} and m_{body} are body height (in m) and weight (kg), respectively, and $c = 36\,\mathrm{kg\,m^{-3}}$ is an empirical constant. Age-dependent h_{body} and m_{body} (GBE 2017) are processed into age-cohort averages (Table 7).

The correction factor, f_{age} (dimensionless), considers variations of the basal metabolic rates, M_{rest}, with age. f_{age} is determined from Altman and Dittmer (1968) data (as reproduced by Cooke 2001) as the ratio of the mean age-cohort M_{rest} to the basal metabolic rate for adults (Table 7). While no TUS activity profiles are available for infants, metabolic rates for appropriate activities are provided (Table B3) and can be paired with the metabolic rates of caretakers.

5.2.2 | Appliance Energy Consumption

For each activity carried out in a residential setting (i.e., 'Home', 'OtherHome'; Figure 1a), the activity profiles give the power of electrical appliances likely used as part of the activity at each timestep (0 W if no appliances are used). Values provided are for actively used appliances only. Thus, energy consumption from

TABLE 7 | Age-cohort dependent parameter values used in Equation (7), derived from averages of male and female samples. Input data sources given in the text.

Age cohort	A_{body} (m^2)	f_{age} (–)
Infant (≤ 4)	0.58	1.32
Child (5–11)	1.06	1.24
Teenager (12–18)	1.67	1.12
Adult (19–64)	1.93	1.00
Senior (≥ 65)	1.90	0.92

devices that run constantly (e.g., refrigerators, WiFi router, and other standby power) or are associated with lighting, heating, ventilation, and air conditioning systems (HVAC) are not included (Section SM5.3).

The appliance power represents the sum of electrical energy consumption associated with all unique TUS activities (up to four) that took place at the same timestep (e.g., primary activity: 'Food preparation and baking', secondary activity: 'Other specified TV watching'). Reference values use current (12/2023) manufacturer power ratings by appliance type (Table B4).

Further to that, for each activity, the likely use of water or lighting in the 'Home' or 'OtherHome' MEs is assessed. This is based solely on the nature of the activity, without considering other factors (e.g., time of day, season, meteorology), which can be accounted for during (e.g., agent-based) modelling. Note that water use is indicated only for activities resulting in open water surfaces (e.g., sink dishwashing, personal hygiene) to enable modelling the impact on turbulent latent heat flux, but not for 'closed units' (e.g., use of washing machines, dishwashers).

6 | Transport

Modelling the movement of citizens between microenvironments of activity (e.g., 'Home' to 'Workplace') located in different parts of the city requires travel route options for different modes of transportation between start and end locations (Figure 1f). A catalogue of available transport routes between all 500-m grid-cell neighbourhoods in London is created by evaluating travel options based on public transport timetables, road network, and speed limits (details in SM6–7). The travel database (Section 6.1) provides viable route stages between neighbourhoods by determining connections available in the local transport network (road and rail). This dataset can then be used by route-finding algorithms that consider aspects of travel duration, cost or service schedule. The latter can be assessed from the capacities of public transport by mode, day type, and time of day provided with the dataset (Section 6.1.3). This allows occupancy of transport microenvironments to be assessed in time and space. To assign travel destinations within Greater London, spatial attractors are defined between all neighbourhoods that measure their attractiveness for all non-'Home' MEs relative to the origin neighbourhood (Sections 6.2 and SM8).

6.1 | Travel Database

The travel database contains collections of route stages from an origin, representing possible connections to other neighbourhoods by different modes of transport: (i) private (driving in private vehicle, walking, bicycling; Section 6.1.1) and (ii) public (bus, tube/subway, train; Section 6.1.2). Each travel stage is a single, uninterrupted route segment that can be completed by the transport modes available, logging stage destination grid, travel time to the destination at 1-min resolution, and road type (for driving, cycling, walking). For public transport, destination grids must be accessible from the origin grid within a single stage by one of the mode options available. For private transport by car or bicycle, the destination grids accessible are a collection of up to eight neighbouring grids if connected by roads, while for walking, all neighbour grids are deemed accessible. The complete route connecting start and end grids is the sequence of individual route stages available from the travel database, which can reflect different user choices, e.g., regarding preferred mode or travel durations.

6.1.1 | Private Transport

The London road network is assessed using Ordnance Survey data (OS Open Roads 2021), which are converted into total road lengths accumulated at the grid neighbourhood scale for four road types (DfT UK 2012): motorways, A roads, B roads, and local roads (Section SM7.3). Road speed limits in each neighbourhood are assigned to each road type by mapping OSM (2020) data or, if missing, national speed limits (GOV UK 2024). Bicycling speeds are fixed to 17 $\mathrm{km\,h^{-1}}$ on local roads and 22 $\mathrm{km\,h^{-1}}$ on A and B roads (motorways are not accessible by bicycle); the walking speed is set to 3 $\mathrm{km\,h^{-1}}$.

Private transport stages for each origin grid are created by assessing available road types and their speed limits in the eight neighbouring grids. For travel by car or bicycle, grids are only considered accessible if the length of at least one road type exceeds 10 m. The time needed to complete the stage is calculated from the mode travel speed and the grid-cell diagonal distance (~700 m).

6.1.2 | Public Transport

Travel options for public transport combine Transport for London timetables (TfL 2022a; bus, tube/subway, train) are extracted for a Monday (for a typical weekday schedule), Saturday, and Sunday in a sample week (7th–14th February 2022) to capture the schedule variability across day types. The public transport database contains information on all possible stages between origin and destination grid cells at all times. Hence, temporal variations of service availability are not directly reflected in the travel database but can be derived in combination with the capacity dataset (Section 6.1.3). This division of service routing and timing allows for different weeks to be sampled according to data availability and intention of modelling (e.g., testing the impact of seasonality or timetable changes during major events or strikes).

An origin grid cell can have multiple public transport routes to many destination neighbourhoods. However, for a given public transport mode (bus, train, tube/subway), origin and destination must both be accessible. This is assessed by identifying public transport access points (e.g., train/tube stations, bus stops) in each grid cell from the TfL (2022a) data. Route stages from the origin grid are created from all unique trips provided by services of any mode that frequent any station within the origin and destination neighbourhoods. The travel duration of a stage is obtained from the difference between the departure times in the origin and destination grid cells, taken from the trip with the largest number of stops on one line. Here a line represents the collection of stops between the start and end points of a trip, which can vary diurnally and between days. This is to ensure that potential differences in the stops frequented by the same service on different trips are considered (e.g., the same tube service can have different start and end stops on different trips throughout the day).

6.1.3 | Public Transport Capacities

To account for variations in the availability of public transport modes based on type of day (Monday, Saturday, Sunday) and time of day (in 10-min time increments, i.e., 144 timesteps in 24 h), the public transport capacity is assessed for all mode types (bus, tube, train). The number of unique trips across all services by mode type leaving each grid cell within each 10-min time window is determined from the TfL (2022a) timetables, and a typical capacity (i.e., number of passengers) of a single trip by mode type is assumed (tube: 796–average of London underground trains, TfL 2022b; train: 811–average of London overground trains, London GOV 2020; bus: 60–most common single-decker bus in London, TfL 2020). The number of departures within a grid-cell neighbourhood is used to calculate a total capacity of the number of people able to use these services at each 10-min timestep. If no services are available in the origin grid at a given time, the capacity is zero.

6.2 | Spatial Attractors

Travel destinations for a change of microenvironment (ME) reflect the availability of certain settings, venues, or natural features in a neighbourhood (e.g., businesses, services, educational settings, parks). To allow a probabilistic but data-driven assignment of the destination in the modelling of travel activities, spatial attractors (or gravity weights) are created that represent the physical and socio-economic context of each grid-cell neighbourhood (Table 8). For any non-'Home' ME other than 'Workplace', the spatial attractor, $\Gamma_{o,d}^{\mathrm{ME}}$, between an origin neighbourhood, o, and a destination, d , is defined as

$$\Gamma_{o,d}^{\mathrm{ME}} = \frac{\gamma_d^{\mathbf{ME}}}{\left(\frac{\delta_{o,d}}{\Delta_N}\right)^2} f_{c,d}^{\mathrm{ME}} \tag{8}$$

where γ is an attractor for the destination grid cell that represents a measure of the 'attractiveness' of the neighbourhood for a given ME. The attractor is scaled by the spatial separation between origin and destination grid cells through the non-dimensional ratio of straight-line distance, δ, and characteristic spatial length scale of the neighbourhood, Δ_N, (here: the 500-m grid length). If the origin is equal to the destination in Equation (8), $\delta_{o,d}$ is set to Δ_N. The attractor, γ, represents a relatively static measure, subject to infrequent changes (e.g., locations of schools, public greenspace, shops; list of attractor types in Table 8).

Time-dependency (e.g., diurnal or seasonal variations) and other factors (e.g., mix of settings, public transport connections, weather conditions) can alter the attractiveness of a neighbourhood which can be considered by defining a community factor, $f_{c,d}^{\mathrm{ME}}$ (Equation 8), set to 1 in the published data.

Spatial attractors for the 'Workplace' ME are derived directly from census data (ONS 2011d) that provide relations between

TABLE 8 | Overview of spatial attractors of different microenvironments characterising each grid-cell neighbourhood.

ME	γ	Description	Derived from
OtherHome	RPop	Residential population	Urban populations (Section 4.1)
University	f_{uni}	Fraction of footprint area of tertiary education buildings to grid area	OS MasterMap (2022)
PrimarySchool	Count	Primary school pupils	London Datastore (2016)
SecondarySchool	Count	Secondary school pupils	London Datastore (2016)
LargeShop	Count	Large commercial/retail	OS POIS (2023)
SmallShop	Count	Small to medium commercial/retail/services	OS POIS (2023)
Hospitality	Count	Restaurant, café, bar, etc.	OS POIS (2023)
Outside	f_{grec}	Fraction of recreational greenspace	Land-cover dataset (Section 2)
OutdoorEnt	Count	Outdoor entertainment venues	OS POIS (2023)
IndoorEnt	Count	Indoor entertainment venues	OS POIS (2023)
Cultural	Count	Indoor leisure, cultural, social venues	OS POIS (2023)
Hotel	Count	Hotel, hostel, guesthouse, etc.	OS POIS (2023)
Healthcare	Count	Hospital, doctor's surgery, health facility, etc.	OS POIS (2023)
TempRes	Count	Temporary residence (e.g., nursing home, prison)	OS POIS (2023)
Workplace	Count	Population working in grid cell by residence grid cell	ONS (2011d)

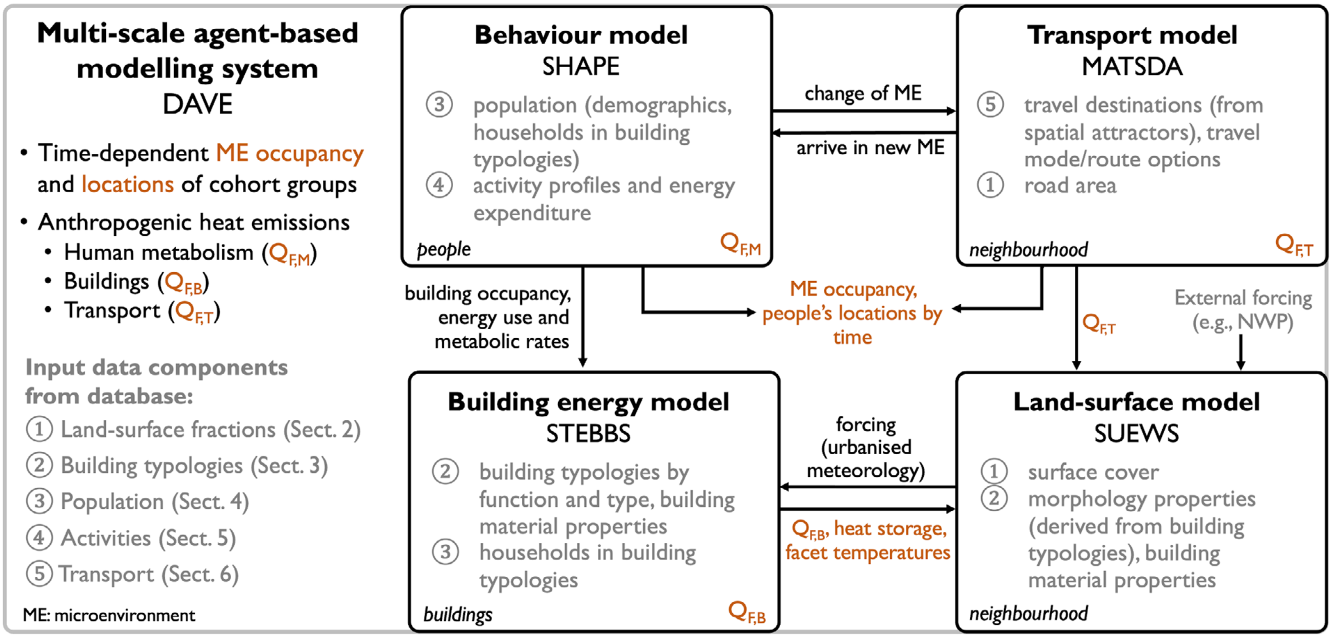

FIGURE 3 | MAPSECC components (grey) used in different sub-modules of the multi-scale agent-based modelling system DAVE (output: orange). Scales (people, building, neighbourhood) of the modules are indicated (bottom left of boxes).

the location of usual residence (OA-level) and the place of work (workplace zone level; WPZ). Mapping of the input data to the grid-cell neighbourhoods is in proportion to the areal overlap between respective OA and WPZ census regions (ONS 2011a) and the grid cell.

7 | Examples of Data

The use of the MAPSECC: London database is exemplified in Figure 3 in the context of the multi-scale agent-based modelling system DAVE (Dynamic Anthropogenic actiVities and feedback to Emissions; Hertwig et al. 2025), with visualisations of some database elements in Figures 4–7. DAVE uses four sub-modules that operate at different scales (people, building, neighbourhood; Figure 3) covered by MAPSECC (Figure 1) to allow simulations of a metropolitan region: (i) a land-surface model (SUEWS, Surface Urban Energy and Water Balance Scheme, e.g., Järvi, Grimmond, and Christen 2011; Ward et al. 2016; Grimmond et al. 2024), (ii) a building energy model (STEBBS, Simplified Thermal Energy Balance for Building Scheme; Capel-Timms et al. 2020), (iii) a behaviour model (SHAPE, Scheduler for Human Activities and energy exPEnditure), and (iv) a transport model (MATSDA, Movement And Transport Simulations using Dijkstra's Algorithm). DAVE provides an assessment of people's location in time (using the 10 min timestep of the activity profiles; Section 5), space (neighbourhoods), and microenvironment (Section 1.1, Table 2) and estimates components of anthropogenic heat emissions.

The land-surface model uses surface-cover fractions (Section 2; Figure 4) and grid-scale morphology properties (Section 3.1.5) to simulate meteorological forcing conditions for the building energy model that reflect the local impact of urban surfaces on the micro-climate (e.g., near-surface temperatures, radiation). The high spatial (500 m) and type resolution of the land-cover database captures the variability of impervious (Figure 4a,b) and natural (Figure 4c–f) surfaces across the city and the diversity of sub-grid scale compositions of land-cover types. Availability of recreational and residential greenspace in neighbourhoods (Figure 4g,h) can be connected to outdoor behaviour in MEs and residential sub-spaces (garden; Table 2a).

The behaviour model uses residential population information with detailed age cohort and household size structure provided for different building typologies (Section 3.1.5) in the assignment of activity profiles (Section 5) and subsequent parsing of building occupancy levels and energy use to the building energy model (BEM). Figure 5 shows the consistency between downscaled residential (Figure 5a) and workplace populations (Figure 5b) and the available total volume of the residential and non-residential building stock identified in each neighbourhood (Figure 5c,d). The building typologies provided to the BEM represent a segmentation of the buildings in each neighbourhood by function and type, with links to typical building materials via the built-period (Sections 3.1.5, 3.2). Compared to national-level building typologies and surveys (e.g., TABULA 2016; English Housing Survey, MHCLG 2024), the database captures the variability of the actual local building stock within and across the London neighbourhoods. Figure 6 illustrates this for residential buildings, with clear changes in the prevalence of specific building types in different regions of the city.

The way the high-resolution input data are processed results in consistency across the different MAPSECC datasets. For example, a small total residential building volume and residential population (Figure 5a,c) occurs in regions primarily characterised by detached houses along the outskirts of the city (Figure 6c). Availability of residential greenspace (e.g., private gardens) is low in the city

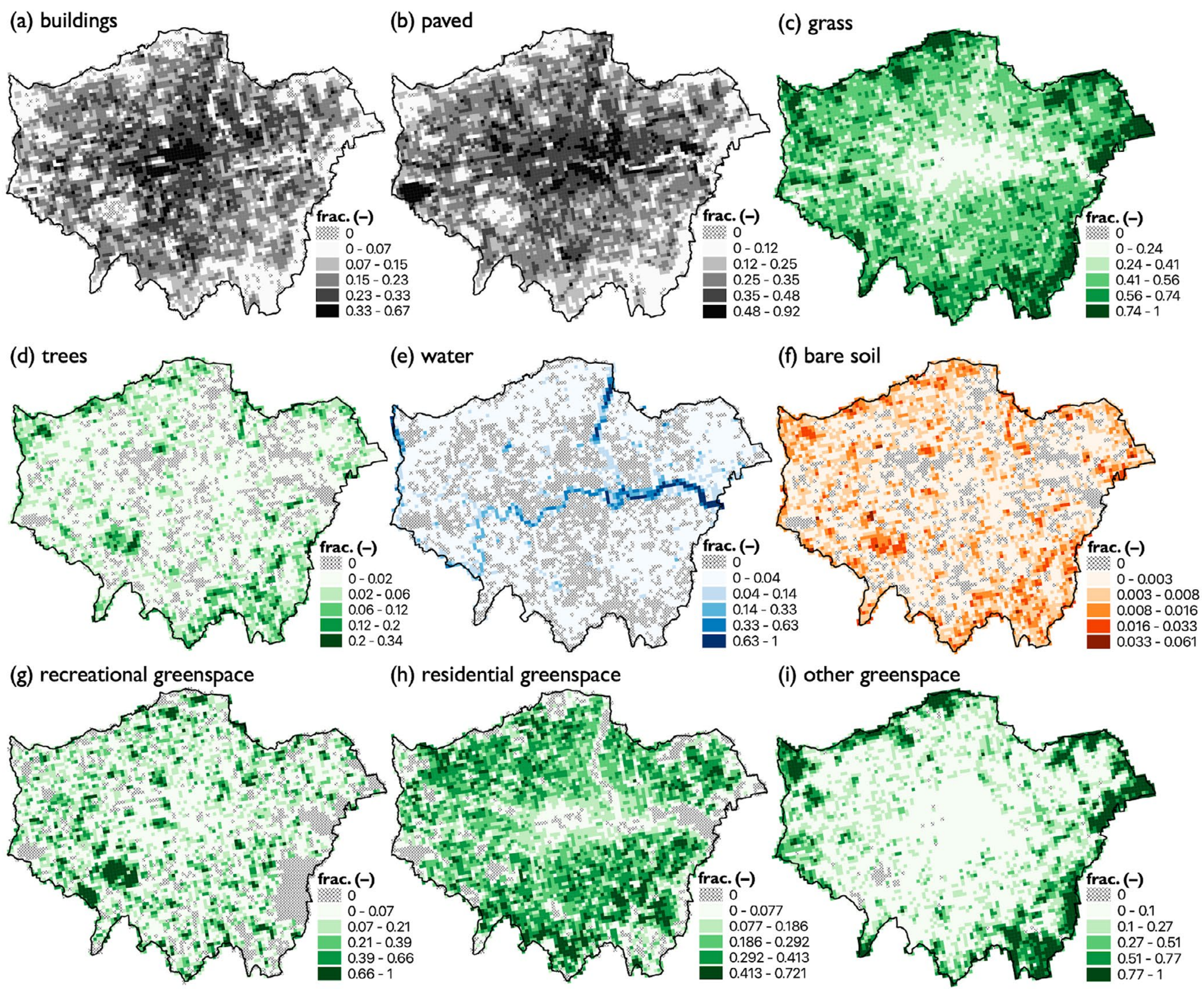

FIGURE 4 | Land-cover fractions in 500 m grid cells for Greater London (black outline) derived from Geomni UKMap (2021) data (Section 2): (a) buildings, (b) paved, (c) grass, (d) trees (combined evergreen and deciduous species fractions), (e) water, (f) bare soil, (g) recreational greenspace, (h) residential (private) greenspace, (i) all remaining greenspace (included neither in g nor h). Data bins are determined from Jenks' (1967) optimisation method. Bin ranges shown are exclusive of the start value and inclusive of the end value, i.e., (start, end].

centre (Figure 5h), where apartment blocks and other types of residential complexes occur more frequently (Figure 6d,e) compared to detached, semi-detached or terraced houses. Small residential building volumes and populations correlate well with land-cover characteristics, e.g., Heathrow Airport (Figure 4b, paved), water surfaces along the rivers Thames and Lee (Figure 4e), and large, non-residential greenspace areas (Figure 4g,i). Thermal and radiative (e.g., heat capacity, albedo, emissivity) building surface properties used in the land-surface modelling can be derived from the effective parameters provided to the BEM (Section 3.2) to ensure consistency between the coupled models.

The simulation of the location of residents throughout the day and the occupancy of MEs uses TUS-derived activity profiles (Section 5). Figure 7a–c shows typical diurnal ME occupancy cycles by aggregating individual profiles from the database. Taking the example of the adult and child age cohorts (Figure 7a,b), the variability of occupancy of residential ('Home', 'OtherHome') and occupational/routine ('Workplace' or 'PrimarySchool') MEs captures differences in the work/school patterns and typical timings of ME changes (e.g., school start/end, Figure 7b). Variations in ME occupancy schedules and amplitudes by type of day are also clearly captured in the profiles of non-routine/leisure MEs, such as 'Hospitality' and other outdoor and indoor settings (Figure 7c).

For the residential 'Home' ME, occupancy of indoor and outdoor spaces in which the primary activities likely have occurred show distinct temporal responses (Figure 7d), e.g., occupancy timings for kitchen (breakfast, lunch, dinner), bathroom (personal hygiene in the mornings, evenings), bedroom (night-time) or garden (daylight hours). The data also capture variations of room occupancy by day type, e.g., shift of the timing of bathroom and kitchen use to later in the mornings as people stay in bed longer on non-workdays *cf.* workdays.

If a change of ME is determined, the transport model uses the spatial attractors (Section 6.2) to determine the location (i.e.,

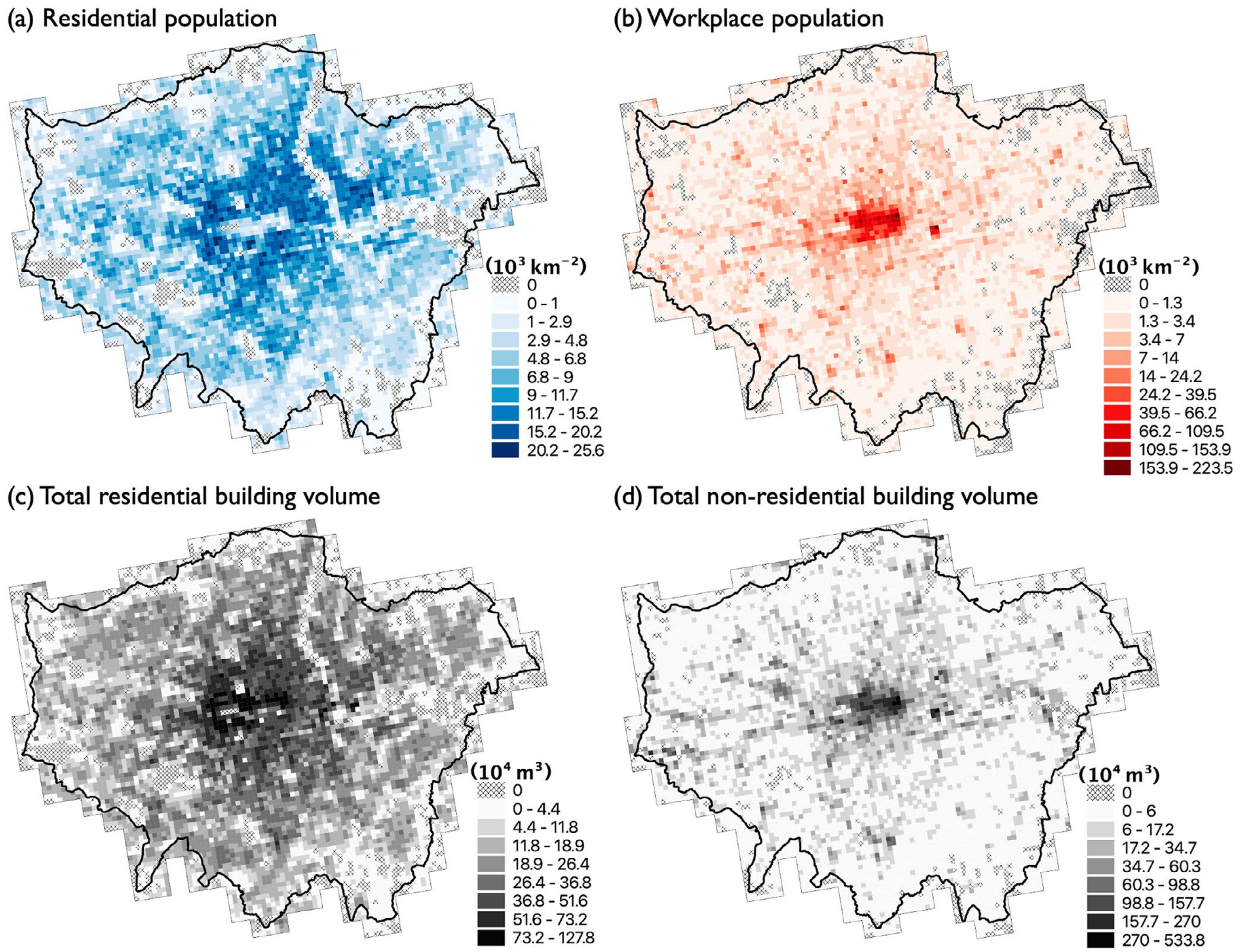

FIGURE 5 | Gridded population and building volume information in 500 m grid cells for Greater London (black outline) derived from population downscaling and building stock analysis (Section 3, Section 4): (a) residential population density, (b) workplace population density, (c) total volume of residential and mixed-use (residential part only) buildings, (d) total volume of non-residential, mixed-use (non-residential part only), and other uninhabitable buildings (Section 3.1.2). The mixed-use building volume is split in equal parts amongst the residential and non-residential portions. Data bins method as in Figure 4.

grid-cell neighbourhood) of the new ME and combines the travel stages available from the transport database (Section 6.1) to a full route using a pathfinding algorithm. This determines people's locations within the city at each model timestep. As the locations of the 'Workplace' ME are determined by the model, the mapped workplace populations (Figure 6b) can be used to evaluate the model results.

8 | Conclusions

A comprehensive database and methodology (MAPSECC) is presented, facilitating dynamic and multi-scale urban modelling that connects physical characteristics of a city (e.g., building geometry, construction materials, surface cover) with socio-economic aspects (building function, urban transport infrastructure, residential and workplace populations, human activity patterns). This uses a methodology that links and harmonises available data resources (Table A1) through spatial downscaling and purpose-specific processing (SM1–8). While the MAPSECC: London database has a specific application as input data to DAVE (Figure 3), all components of the database have wider usability beyond this purpose:

- *Land-cover fractions*: A description of surface-cover composition is essential for modelling the urban surface energy balance and with it the representation of micro-meteorological features of the urban climate. The relatively high resolution of the gridded dataset (500 m) captures spatial variations of the sub-grid scale land-cover mix between neighbourhoods and across the city from the densely built-up centre to the greener suburbs (Figure 4). This allows local differences of the urban micro-climate to be represented.
- *Building typologies*: Local building typologies with morphology and population information for different building functions, types, and construction materials (linked via building age bands) in > 7,000 grid-cell neighbourhoods are derived from a data-enriched building-level database. Building typologies reflect the spatial variability and diversity of the city's building stock within and across neighbourhoods and enable city-wide modelling of the

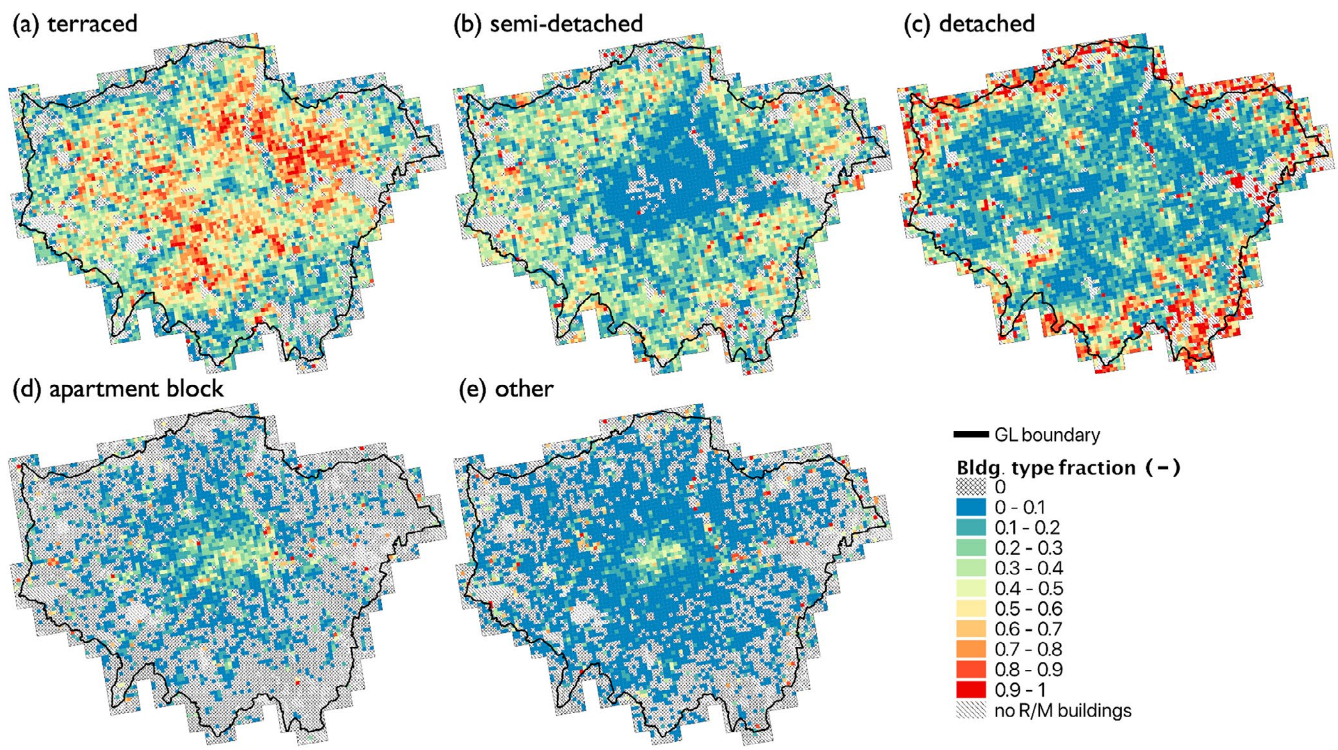

FIGURE 6 | Ratio of the building volume of residential (R) and mixed-use (M) buildings by building type (Section 3.1.3) to the total residential and mixed-use building volume (Figure 5c) derived in 500 m grid cells covering Greater London. Only the residential use part of the mixed-use building volume is accounted for (assuming a 50% split). Bin range characteristics as in Figure 4.

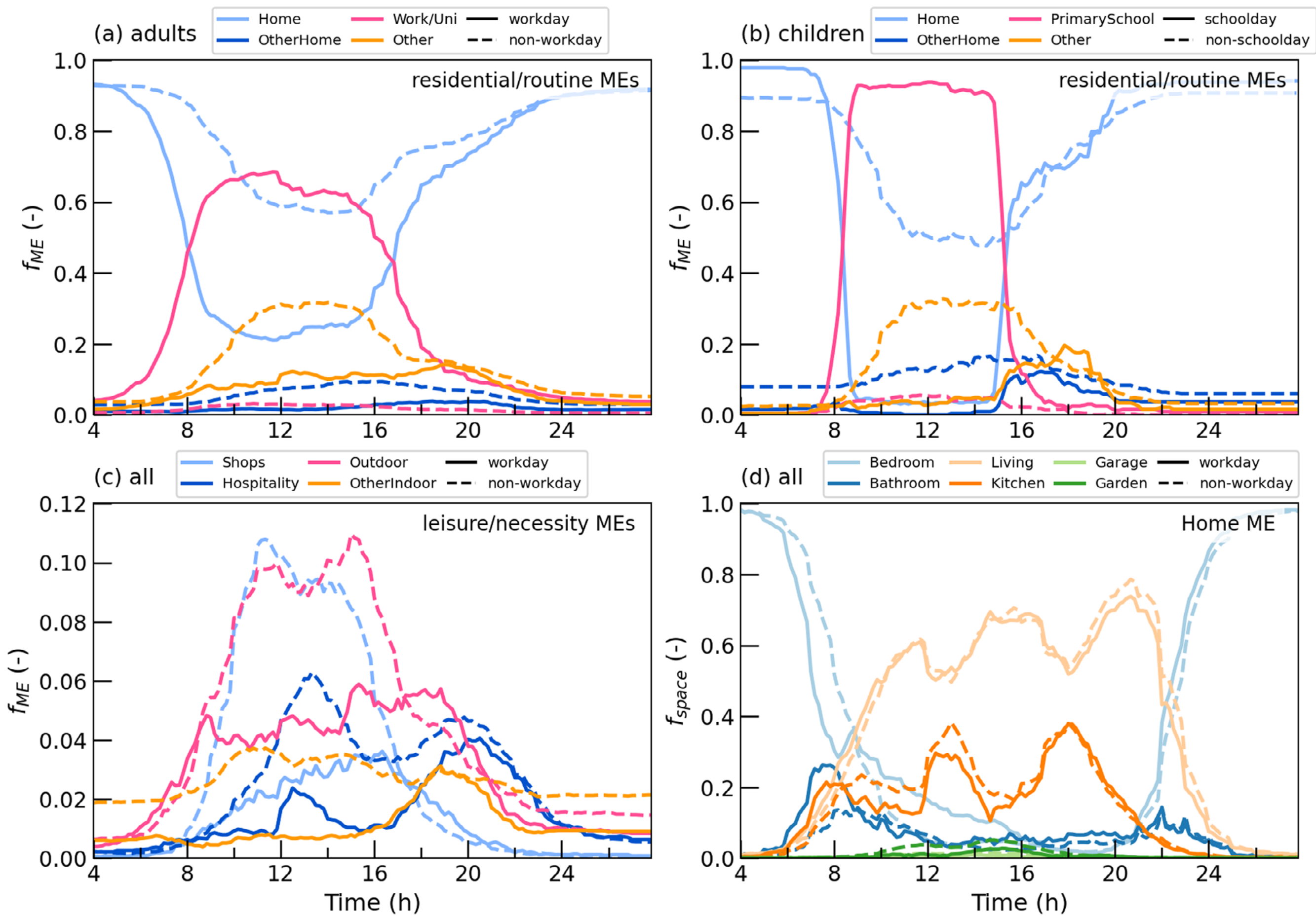

FIGURE 7 | Legend on next page.

FIGURE 7 | Fractional diurnal (4 am to 4 am) occupancies on (non-)work/schooldays from UK–TUS 2014–15 derived activity profiles (Section 5) of (a) adults (age 19–64) and (b) children (age 8–11) in residential and routine MEs; (c) all age groups in MEs of necessity and leisure activities ('Shops' combines 'SmallShop', 'LargeShop'; 'Outdoor' groups 'Outside', 'RoundTrip', 'OutdoorEnt'; remaining non-residential/non-routine indoor MEs are grouped into 'OtherIndoor'); (d) all age groups in indoor and outdoor sub-spaces of the 'Home' ME (including the 'Sleep' ME state).

thermal performance of buildings with urban building energy models, e.g., to study anthropogenic heat emissions or indoor thermal comfort.

- *Human activity profiles and energy expenditure*: The novel analysis of time-use survey data creates links to a variety of indoor and outdoor microenvironments of activity that are relevant for heat stress and air-pollution exposure studies. Such assessments benefit from the day type, season, and age-cohort dependent variations of occupancy of microenvironments and activities captured. The combination of activity- and time-dependent metabolic rates, energy consumption behaviours and room occupancy (needed to determine shared energy use of appliances and lighting within a household) facilitate more realistic building energy modelling in residential settings, as demonstrated by Liu et al. (2024).
- *Transport routes*: The database of public and private transport options across Greater London allows the modelling of realistic transport routes between neighbourhoods by mode type and with it the occupancy in transport microenvironments. This supports the modelling of anthropogenic heat emissions from traffic sources and can be equally useful for air-pollution exposure assessments outdoors and in vehicles.

Where possible, in generating the MAPSECC: London database we have used data from government agencies instead of relying more on crowd-sourced data (e.g., OpenStreetMap), assuming a robust quality control and consistency between data releases, even though update cycles may be less rapid. The MAPSECC processing is intentionally repeatable for other UK cities, as it relies to a large degree on primary input data available at the national level (Office of National Statistics, Ordnance Survey, Valuation Office Agency; Table A1). However, adjustments (e.g., suitable code interfaces; SM1–8) are needed to deal with region-specific information and data sources (e.g., public transport systems and capacities; school populations). As building material parameters and activity profiles use national data, they are applicable across the UK.

The downscaling of information to the building-level allowed for a flexible and fit-for-purpose spatial aggregation of data that is not confined by census or administrative boundaries. This required very high-resolution input information that may not be available in other countries, but moving to coarser resolutions in principle is possible (tolerating larger uncertainties in derived information). Analyses are currently undertaken to determine what data compromises are acceptable for which type of modelling application and research foci (processes), with initial analyses in Berlin (Fenner et al. 2024) being followed by other cities (e.g., Bristol, Paris).

The MAPSECC: London data repository will be updated continuously as (i) new or updated primary input data become available, (ii) enhancements of existing database components occur and (iii) new data components are added. As previous database versions continue to be available, this allows inter-comparison between different states, e.g., to determine temporal changes in urban populations and their activities reflected in census or time use survey data. We hope that such enhancements, including the potential expansion of the current scope of MAPSECC, will also be guided by feedback from users of the database.

Acknowledgements

We gratefully acknowledge funding by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 855005; project *urbisphere*), the UKRI Natural Environment Research Council (NERC) projects APEx (NE/T001887/2), and ASSURE (NE/W002965/1), and the UKRI Engineering and Physical Sciences Research Council (EPSRC) project CREDS (EP/R035288/1). We thank our colleagues for helpful discussions and feedback on the database. We are grateful to the reviewers for their enthusiasm for the database and for pointing out potential for its expansion.

Data Availability Statement

The database is published on Zenodo (https://zenodo.org/records/12190340) with an embargo period ending on 1st April 2027. Requests for access prior to this date are enabled.

References

Ainsworth, B., W. L. Haskell, A. S. Leon, et al. 1993. "Compendium of Physical Activities, Classification of Energy Costs in Human Physical Activities." *Medicine and Science in Sports and Exercise* 25: 71–80. https://doi.org/10.1249/00005768-199301000-00011.

Altman, P. L., and D. S. Dittmer. 1968. *Metabolism*. Bethesda, MD: Federation of American Societies for Experimental Biology.

ASHRAE. 2017. *ANSI/ASHRAE Standard 55–2017. Thermal Environmental Conditions for Human Occupancy*. Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

BS EN ISO 8996. 2004. *Ergonomics of the Thermal Environment – Determination of Metabolic Heat Production*. Geneva, Switzerland: International Organization for Standardization.

Capel-Timms, I., S. T. Smith, T. Sun, and S. Grimmond. 2020. "Dynamic Anthropogenic activitieS Impacting Heat Emissions (DASH v1.0): Development and Evaluation." *Geoscientific Model Development* 13: 4891–4924. https://doi.org/10.5194/gmd-13-4891-2020.

Ching, J., D. Aliaga, G. Mills, et al. 2019. "Pathway Using WUDAPT's Digital Synthetic City Tool Towards Generating Urban Canopy Parameters for Multi-Scale Urban Atmospheric Modeling." *Urban Climate* 28: 100459. https://doi.org/10.1016/j.uclim.2019.100459.

CIBSE. 2015. *Guide A: Environmental Design*. London, UK: Chartered Institution of Building Services Engineers. https://www.cibse.org/knowledge-research/knowledge-portal/guide-a-environmental-design-2015.

Cooke, C. B. 2001. "Metabolic Rate and Energy Balance. Chapter 6." In *Kinanthropometry and Exercise Physiology Laboratory Manual: Tests, Procedures and Data, Volume Two: Exercise Physiology*, edited by

R. Eston and T. Reilly, 2nd ed., 161–188. London, UK; New York, NY: Routledge.

Demuzere, M., S. Hankey, G. Mills, W. Zhang, T. Lu, and B. Bechtel. 2020. "Combining Expert and Crowd-Sourced Training Data to Map Urban Form and Functions for the Continental US." *Scientific Data* 7: 264. https://doi.org/10.1038/s41597-020-00605-z.

Department for Transport, UK (DfT UK). 2012. "Statutory Guidance: Guidance on Road Classification and the Primary Route Network." Accessed May 30, 2024. https://www.gov.uk/government/publications/guidance-on-road-classification-and-the-primary-route-network/guidance-on-road-classification-and-the-primary-route-network.

DOE. 2024. "EnergyPlus™ Version 24.1.0 Documentation, Engineering Reference (Build: 9d7789a3ac)." U.S. Department of Energy. Accessed May 30, 2024. https://energyplus.net/.

Duan, N. 1982. "Models for Human Exposure to Air Pollution." *Environment International* 8, no. 1–6: 305–309. https://doi.org/10.1016/0160-4120(82)90041-1.

Fenner, D., A. Christen, S. Grimmond, et al. 2024. "Urbisphere-Berlin Campaign: Investigating Multiscale Urban Impacts on the Atmospheric Boundary Layer." *Bulletin of the American Meteorological Society* 105: E1929–E1961. https://doi.org/10.1175/BAMS-D-23-0030.1.

Gastineau, P., A. Can, B. F. Yaméogo, et al. 2023. "Modeling Exposure to Mobility-Related Pollution: Review and Key Challenges." *Transportation Research Record* 2678: 587–603. https://doi.org/10.1177/03611981231208907.

GBE. 2017. "Height and Weight of the Population, 2017 (Table, Ad-Hoc)." In *The Federal Health Monitoring System (GBE), Germany, Based on Microcensus Health Questions*. Bonn: Federal Statistical Office. Accessed March 1, 2024. www.gbe-bund.de.

Geomni UKMap. 2021. "UKMap Edition 26 [FileGeoDatabase Geospatial Data], Scale 1:1000, Tiles: GB, Updated: 1 March 2021, Verisk/Geomni." Accessed Using: EDINA Geomni Digimap Service. Accessed July 7, 2021. https://digimap.edina.ac.uk.

Gershuny, J., and O. Sullivan. 2017. "United Kingdom Time Use Survey, 2014–2015." Centre for Time Use Research, IOE, University College London [Data Collection]. UK Data Service. SN: 8128. https://doi.org/10.5255/UKDA-SN-8128-1.

GOV UK. 2024. "National Speed Limits." Accessed May 3, 2024. https://www.gov.uk/speed-limits.

Grimmond, S., V. Bouchet, L. T. Molina, et al. 2020. "Integrated Urban Hydrometeorological, Climate and Environmental Services: Concept, Methodology and Key Messages." *Urban Climate* 33: 100623. https://doi.org/10.1016/j.uclim.2020.100623.

Grimmond, S., L. Järvi, H. Ward, et al. 2024. "SUEWS (Surface Urban Energy and Water Balance Scheme) (Latest-Dev)." *Zenodo*. https://doi.org/10.5281/zenodo.11164910.

Hertwig, D., M. McGrory, Y. Liu, et al. 2025. "DAVE (Dynamic Anthropogenic actiVities and feedback to Emissions) Documentation and Manual (2025_1.0)." Zenodo. https://doi.org/10.5281/zenodo.11369893.

International Code Council (ICC). 2021. *2021 International Energy Conservation Code*. Washington, DC: International Code Council. https://codes.iccsafe.org/content/IECC2021P1.

Järvi, L., C. S. B. Grimmond, and A. Christen. 2011. "The Surface Urban Energy and Water Balance Scheme (SUEWS): Evaluation in Los Angeles and Vancouver." *Journal of Hydrology* 411, no. 3–4: 219–237. https://doi.org/10.1016/j.jhydrol.2011.10.001.

Jenks, G. F. 1967. "The Data Model Concept in Statistical Mapping." *International Yearbook of Cartography* 7: 186–190.

Jia, S., and Y. Wang. 2021. "Effect of Heat Mitigation Strategies on Thermal Environment, Thermal Comfort, and Walkability: A Case Study in Hong Kong." *Building and Environment* 201: 107988. https://doi.org/10.1016/j.buildenv.2021.107988.

Lipson, M. J., S. Grimmond, M. Best, 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, no. 758: 126–169. https://doi.org/10.1002/qj.4589.

Liu, Y., S. Grimmond, Z. Luo, et al. 2024. "Dwelling Conversion and Energy Retrofit Modify Building Anthropogenic Heat Emission Under Past and Future Climates: A Case Study of London Terraced Houses." *Energy and Buildings* 322: 114668. https://doi.org/10.1016/j.enbuild.2024.114668.

LoC. 2024a. "US Library of Congress, Sustainability of Digital Formats: ESRI Shapefile." Accessed May 2, 2024. https://www.loc.gov/preservation/digital/formats/fdd/fdd000280.shtml.

LoC. 2024b. "US Library of Congress, Sustainability of Digital Formats: CSV, Comma Separated Values (RFC 4180)." Accessed May 2, 2024. https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml.

LoC. 2024c. "US Library of Congress, Sustainability of Digital Formats: JSON (JavaScript Object Notation)." Accessed May 2, 2024. https://www.loc.gov/preservation/digital/formats/fdd/fdd000381.shtml.

Loga, T., B. Stein, and N. Diefenbach. 2016. "TABULA Building Typologies in 20 European Countries—Making Energy-Related Features of Residential Building Stocks Comparable." *Energy and Buildings* 132: 4–12. https://doi.org/10.1016/j.enbuild.2016.06.094.

London Climate Change Partnership (LCCP). 2024. Accessed June 19, 2024. http://climatelondon.org.uk/lccp/.

London Datastore. 2014. "Statistical GIS Boundary Files for London." Accessed March 20, 2024. https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london.

London Datastore. 2016. "London Schools Atlas." Accessed August 12, 2021. https://data.london.gov.uk/dataset/london-schools-atlas.

London GOV. 2020. "London Overground Capacity." Accessed May 29, 2024. https://www.london.gov.uk/who-we-are/what-london-assembly-does/questions-mayor/find-an-answer/london-overground-capacity.

Lund, A. M., R. Gouripeddi, and J. C. Facelli. 2020. "STHAM: An Agent Based Model for Simulating Human Exposure Across High Resolution Spatiotemporal Domains." *Journal of Exposure Science & Environmental Epidemiology* 30: 459–468. https://doi.org/10.1038/s41370-020-0216-4.

Masson, V., W. Heldens, E. Bocher, et al. 2020. "City-Descriptive Input Data for Urban Climate Models: Model Requirements, Data Sources and Challenges." *Urban Climate* 31: 100536. https://doi.org/10.1016/j.uclim.2019.100536.

MHURD. 2015. "Ministry of Housing and Urban-Rural Development of the People's Republic of China." Design Standard for Energy Efficiency of Public Buildings, GB 50189-2015, Beijing, China.

Ministry of Housing, Communities and Local Government (MHCLG). 2004. "Age of Commercial and Industrial Stock, Borough. Creator: Department for Communities and Local Government (DCLG)." Accessed December 3, 2020. https://data.london.gov.uk/dataset/age-commercial-and-industrial-stock-borough.

Ministry of Housing, Communities and Local Government (MHCLG). 2024. "English Housing Survey. [Data Series]. 6th Release. UK Data Service." SN: 200010. Accessed May 30, 2024. https://doi.org/10.5255/UKDA-Series-200010.

Mosteller, R. D. 1987. "Simplified Calculation of Body-Surface Area." *New England Journal of Medicine* 317: 1098. https://doi.org/10.1056/NEJM198710223171717.

ONS. 2011a. "Office for National Statistics 2011 Census: Boundary Data (England and Wales) [Data Collection]." UK Data Service.

SN:5819 UKBORDERS: Digitised Boundary Data. Accessed January 21, 2021. http://census.ukdataservice.ac.uk/get-data/boundary-data.aspx. Contains Public Sector Information Licensed Under the Open Government Licence v3. https://borders.ukdataservice.ac.uk/easy_download.html.

ONS. 2011b. “Office for National Statistics 2011: English Census Output Areas, 2011, Clipped and Generalised With Univariate Census 2011 Data [Data Collection].” From GeoLytix UK 2011 Census Pack. Accessed Through UK Data Service. Contains Public Sector Information Licensed Under the Open Government Licence v3. Accessed January 21, 2021. https://borders.ukdataservice.ac.uk/easy_download_data.html?data=England_oa_2011.

ONS. 2011c. “Office for National Statistics 2011: Estimates of the Workplace Population.” Table “WP101EW Population (Workplace Population)”. Accessed March 21, 2022. Accessed from Nomis: https://www.nomisweb.co.uk/query/construct/summary.asp?mode=construct&version=0&dataset=1300.

ONS. 2011d. “Office for National Statistics 2011: Location of Usual Residence and Place of Work (With Outside UK Collapsed) (OA/WPZ Level).” Accessed March 10, 2023. Accessed from Nomis: https://www.nomisweb.co.uk/census/2011/wf02ew.

OS MasterMap. 2019. “Ordnance Survey MasterMap With Building Height Attribute [FileGeoDatabase Geospatial Data], Scale 1:2500, Data Download Version: October 2019, Ordnance Survey (GB), Using: EDINA Digimap Ordnance Survey Service.” Accessed December 9, 2020 and March 18, 2021. https://digimap.edina.ac.uk.

OS MasterMap. 2022. “Ordnance Survey MasterMap Sites Layer [SHAPE Geospatial Data], Scale 1:2500, Tiles: GB, Updated: 1 October 2022, Ordnance Survey (GB), Using: EDINA Digimap Ordnance Survey Service.” Accessed August 4, 2023. https://digimap.edina.ac.uk.

OS Open Roads. 2021. “Ordnance Survey Open Roads Dataset.” Version Date: 2021-04. Accessed March 16, 2024. https://osdatahub.os.uk/downloads/open/OpenRoads.

OS POIS. 2023. “Ordnance Survey Points of Interest [FileGeoDatabase Geospatial Data], Scale 1:1250, Tiles: GB, Updated: 1 June 2023, Ordnance Survey (GB), Using: EDINA Digimap Ordnance Survey Service.” Accessed August 4, 2023. https://digimap.edina.ac.uk.

OSM. 2020. “OpenStreetMap Data Files for England, December 2020 Version.” Accessed December 14, 2020. http://download.geofabrik.de/europe/united-kingdom/england.html.

Pfafferott, J., S. Rißmann, M. Sühring, F. Kanani-Sühring, and B. Maronga. 2021. “Building Indoor Model in PALM-4U: Indoor Climate, Energy Demand, and the Interaction Between Buildings and the Urban Microclimate.” *Geoscientific Model Development* 14: 3511–3519. https://doi.org/10.5194/gmd-14-3511-2021.

RBKC. 2014. “Borough of Kensington and Chelsea Council.” Accessed March 21, 2024. https://www.rbkc.gov.uk/pdf/school_term_hol_dates_2013-14.pdf.

RBKC. 2015. “Borough of Kensington and Chelsea Council.” Accessed March 21, 2024. https://www.rbkc.gov.uk/pdf/school_term_holiday_dates_2014-15.pdf.

RBKC. 2016. “Borough of Kensington and Chelsea Council.” Accessed March 14, 2024. https://www.rbkc.gov.uk/pdf/School%20terms%20and%20holiday%20dates%202015-2016.pdf.

Smith, J. D., C. Mitsakou, N. Kitwiroon, et al. 2016. “London Hybrid Exposure Model: Improving Human Exposure Estimates to NO_2 and PM2.5 in an Urban Setting.” *Environmental Science & Technology* 21: 11760–11768. https://doi.org/10.1021/acs.est.6b01817.

TABULA. 2016. “IEE Project TABULA (2009-2012). Typology Approach for Building Stock Energy Assessment.” Accessed March 18, 2024. https://episcope.eu/iee-project/tabula/; https://episcope.eu/building-typology/overview/.

Takane, Y., K. Nakajima, and Y. Kikegawa. 2022. “Urban Climate Changes During the COVID-19 Pandemic: Integration of Urban-Building-Energy Model With Social Big Data.” *npj Climate and Atmospheric Science* 5: 44. https://doi.org/10.1038/s41612-022-00268-0.

TfL. 2020. “All London Buses Now Taking Payments as Capacity Limits Are Safely Increased.” Accessed May 29, 2024. https://tfl.gov.uk/info-for/media/press-releases/2020/july/all-london-buses-now-taking-payments-as-capacity-limits-are-safely-increased.

TfL. 2022a. “Transport for London Journey Planner Timetables Open Data.” (This Always Links to the Latest Version). Accessed February 7, 2022. https://tfl.gov.uk/info-for/open-data-users/our-open-data; https://tfl.gov.uk/tfl/syndication/feeds/journey-planner-timetables.zip.

TfL. 2022b. “London Underground Rolling Stock Information Sheets.” 4th ed. Accessed May 29, 2024. https://tfl.gov.uk/corporate/transparency/freedom-of-information/foi-request-detail?referenceId=FOI-0237-2223; https://foi.tfl.gov.uk/FOI-0237-2223/RS%20Info%20Sheets%204%20Edition.pdf.

Valuation Office Agency (VOA). 2020a. “Official Statistics Council Tax: Stock of Properties, 2020. Table CTSOP3.1: Number of Properties by Council Tax Band, Property Type and Region, County, Local Authority District and Lower and Middle Super Output Area.” Accessed February 24, 2021. https://www.gov.uk/government/statistics/council-tax-stock-of-properties-2020.

Valuation Office Agency (VOA). 2020b. “Official Statistics Council Tax: Stock of Properties, 2020.” Table CTSOP4.1: Number of Properties by Council Tax Band, Property Build Period and Region, County, Local Authority District and Lower and Middle Super Output Area. Accessed February 24, 2021. https://www.gov.uk/government/statistics/council-tax-stock-of-properties-2020.

Ward, H. C., and C. S. B. Grimmond. 2017. “Assessing the Impact of Changes in Surface Cover, Human Behaviour and Climate on Energy Partitioning Across Greater London.” *Landscape and Urban Planning* 165: 142–161. https://doi.org/10.1016/j.landurbplan.2017.04.001.

Ward, H. C., S. Kotthaus, L. Järvi, and C. S. B. Grimmond. 2016. “Surface Urban Energy and Water Balance Scheme (SUEWS): Development and Evaluation at Two UK Sites.” *Urban Climate* 18: 1–32. https://doi.org/10.1016/j.uclim.2016.05.001.

Yang, L. E., P. Hoffmann, J. Scheffran, S. Rühe, J. Fischereit, and I. Gasser. 2018. “An Agent-Based Modeling Framework for Simulating Human Exposure to Environmental Stresses in Urban Areas.” *Urban Science* 2, no. 2: 36. https://doi.org/10.3390/urbansci2020036.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Appendix A

An overview of all primary data sources used in the generation of the MAPSECC: London database (Table 1) is given in Table A1, with further information found in the main text and all technical aspects of the processing presented in detail in SM1–8, including meta information for all datasets (input and output).

TABLE A1 | Input data processed to generate the MAPSECC: London urban modelling database (Table 1). OA: Output Area, LSOA: Lower layer Super Output Area, MSOA: Middle layer Super Output Area, WPZ: Workplace zone.

Input data information	Source	Resolution	Version
Building footprints with height attributes	OS MasterMap (2019)	Vector, polygons	2019/10
Infrastructure and functional site locations	OS MasterMap (2022)	Vector, polygons	2022/10
Location of businesses, services, venues	OS POIS (2023)	Vector, points	2023/06
Road network with road types	OS Open Roads (2021)	Vector, lines	2021/04
English census boundaries	ONS (2011a)	OA, LSOA, MSOA, WPZ	2011
English census data	ONS (2011b)	OA	2011
Workplace population	ONS (2011c)	OA	2011
Locations of residence and workplace	ONS (2011d)	OA, WPZ	2011
London borough boundaries	London Datastore (2014)	Administrative boundaries	2011
London school population	London Datastore (2016)	Vector, points	2016
Number of properties by type	VOA (2020a)	LSOA	2020
Number of properties by built period	VOA (2020b)	LSOA	2020
Age of non-residential buildings	MHCLG (2004)	London borough	2004
OpenStreetMap data for England	OSM (2020)	Vector, polygons, points	2020/12
Land cover/use for Greater London	Geomni UKMap (2021)	Vector, polygons	2021/03
UK Time Use Survey	Gershuny and Sullivan (2017)	10 min	2014/15
School calendars	RBKC (2014, 2015, 2016)	Daily	2014–2016
Metabolic rates	Altman and Dittmer (1968), Mosteller (1987), Ainsworth et al. (1993), Cooke (2001), BS EN ISO 8996 (2004), ASHRAE (2017), GBE (2017)	Person by age group	—
Appliance power ratings	Manufacturer information	Appliance types	2023/12
Public transport timetables	TfL (2022a)	1 min	2022/02/07–14
Public transport capacities	TfL (2020, 2022b), London GOV (2020)	Single trips of tube, train, bus	2020, 2022
Road speed limits (England)	GOV UK (2024)	Road type/function	2024

Appendix B

A summary of sample sizes (unique UK–TUS diaries) stratified by age group and diary date is given in Table B1. Table B2 presents examples of diurnal sequences of activity profiles from the processed dataset. A selection of metabolic rates for infants (by activity type) is given in Table B3, to enable modelling caretaker–infant metabolic output (Section 5.2.1). Power ratings of a selection of electrical appliances used in the mapping of energy consumption to TUS activities (Section 5.2.2) are given in Table B4.

TABLE B1 | Sample sizes of unique UK–TUS 2014–15 diaries included in the activity profile dataset, by age cohort, day type (0: weekday without school; 1: weekend; 2: weekday with school) and season of response (winter–DJF, spring–MAM, summer–JJA, autumn–SON; months of season indicated by first letter).

Age cohort	Day type	DJF	MAM	JJA	SON	All
Child	0	18	22	48	13	914
	1	93	127	104	126	
	2	82	102	56	123	
Teenager	0	17	39	71	23	1515
	1	153	176	182	235	
	2	139	137	122	221	
Adult	0	160	263	488	185	10,590
	1	1211	1263	1223	1538	
	2	1103	996	772	1388	
Senior	0	44	92	147	37	3325
	1	375	405	399	468	
	2	350	307	257	444	

TABLE B2 | Example daily sequences of activity profiles derived from UK–TUS 2014–15 responses of an adult diarist (age cohort 19–64) on a (a) workday and (b) non-workday. Timesteps are 10 min: 144 in total from 4 am (timestep 24) to 3:50 am (timestep 23) on the following day. For clarity, only unique TUS primary activities and residential sub-spaces are shown for each profile, while the full dataset provides all details at each timestep. The final ME of the day will be unchanged through to timestep 23 on the next day (i.e., last diary timestep), so "next ME" at that time is always the same as "current ME."

Start timestep	Local time	Current ME	Next ME	Unique primary activities in current ME	Unique space in 'Home' ME
(a) 24	4:00	Sleep	Home	Sleep	Bedroom
42	7:00	Home	Workplace	Other personal care: Wash and dress	Bathroom
48	8:00	Workplace	Home	Travel to work from home and back only; Main job: Working time in main job	—
108	18:00	Home	Hospitality	Travel to work from home and back only; Eating; Unspecified TV video or DVD watching	Kitchen, living room
120	20:00	Hospitality	Sleep	Other specified social life	—
138	23:00	Sleep	Sleep	Sleep	Bedroom
(b) 24	4:00	Sleep	Home	Sleep	Bedroom
40	6:40	Home	Hospitality	Sleep: In bed not asleep; Food preparation and baking; Eating, Other specified TV watching; Resting—Time out; Other personal care: Wash and dress	Bedroom, kitchen, living room, bathroom
71	11:50	Hospitality	LargeShop	Travel related to other social activities; Other specified social life	—
85	14:10	LargeShop	Home	Travel related to shopping; Unspecified shopping	—
90	15:00	Home	Sleep	Travel related to shopping; Food preparation and baking; Resting—Time out, Eating; Communication on the internet; Dish washing; Unspecified TV video or DVD watching; Other personal care: Wash and dress	Kitchen, living room, bathroom
2	00:20	Sleep	Sleep	Sleep	Bedroom

TABLE B3 | Metabolic rates, M_{act}, for infants (age 0–4) for selected UK–TUS 2014–15 activity types calculated from Equation (7), Table 7 and MET_{act} from Ainsworth et al. (1993) and ASHRAE (2017).

TUS activity match for infants (age ≤4)	MET_{act} (–)	M_{act} (W)
'Sleep'	0.7	30
'Resting – time out'	0.9	38
'Solo games and play'	1.5	63
'Accompanying child'	2.6	110
'Visiting an urban park playground designated play area'	3.5	147
'Unspecified ball games'	7.0	295

TABLE B4 | Power ratings of electrical appliances used in residential MEs ('Home', 'OtherHome') during certain UK–TUS activities. Values are based on manufacturer power ratings of typical devices sold in the UK (12/2023). Data sources and reference sheets for these values are provided in the data repository (further details in SM5).

Electrical appliance type	Power (W)
TV	71
Oven/cooker	810
Iron	2800
Washing machine	530
Gaming console	210
Desktop computer (with monitor)	201
Laptop	70
Vacuum cleaner	620
Small appliances (e.g., radio, phone)	5